

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:46:01; Search time 28.95 Seconds
(without alignments)
285,176 Million cell updates/sec

Title: US-09-980-881-3

Perfect score: 1798
Sequence: 1 FOSGQVLAALPRISROYQVLT.....IKYSFTSNPVEKILPLSLK 338

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCRTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1699.5	94.5	423	1	US-07-649-591B-3
2	1699.5	94.5	423	1	US-08-277-540-3
3	1699.5	94.5	423	1	US-08-430-787A-3
4	1699.5	94.5	423	2	US-08-869-057-2
5	646.5	36.0	404	1	US-08-696-139-2
6	637.5	35.5	415	2	US-08-860-882A-57
7	632	35.2	417	1	US-07-649-591B-7
8	632	35.2	417	1	US-08-277-540-7
9	632	35.2	417	1	US-08-430-787A-7
10	622	34.6	417	1	US-07-649-591B-6
11	622	34.6	417	1	US-08-277-540-6
12	622	34.6	417	1	US-08-430-787A-6
13	619.5	34.5	716	4	US-09-171-945-125
14	610	33.9	307	2	US-08-782-760-6
15	610	33.9	307	5	PCT-US96-00995-6
16	609.5	33.9	396	1	US-07-649-591B-4
17	609.5	33.9	396	1	US-08-277-540-4
18	609.5	33.9	396	1	US-08-430-787A-4
19	582	32.4	306	1	US-08-696-139-4
20	549.5	30.6	417	1	US-07-649-591B-8
21	549.5	30.6	417	1	US-08-277-540-8
22	549.5	30.6	417	1	US-08-430-787A-8
23	549	29.5	613	4	US-09-171-945-113
24	532.5	29.6	419	1	US-07-649-591B-5
25	532.5	29.6	419	1	US-08-277-540-5
26	532.5	29.6	419	1	US-08-430-787A-5
27	516.5	28.7	417	4	US-08-640-906-4

28	516.5	28.7	417	4	US-09-395-936-4	Sequence 4, Appl
29	515.5	28.7	417	4	US-08-640-906-18	Sequence 18, Appl
30	515.5	28.7	417	4	US-09-395-936-18	Sequence 18, Appl
31	491.5	27.3	419	4	US-08-640-906-2	Sequence 2, Appl
32	491.5	27.3	419	4	US-09-395-936-2	Sequence 2, Appl
33	486.5	27.1	419	4	US-08-640-906-17	Sequence 17, Appl
34	486.5	27.1	419	4	US-09-395-936-17	Sequence 17, Appl
35	167	9.3	37	1	US-07-649-591B-1	Sequence 1, Appl
36	167	9.3	37	1	US-08-277-540-1	Sequence 1, Appl
37	167	9.3	37	1	US-08-430-787A-1	Sequence 1, Appl
38	116.5	6.5	706	1	US-08-484-105-16	Sequence 16, Appl
39	116.5	6.5	706	1	US-08-484-106-16	Sequence 16, Appl
40	102	5.7	734	4	US-08-706-216-2	Sequence 2, Appl
41	97	5.4	95	2	US-08-782-760-3	Sequence 2, Appl
42	97	5.4	95	5	PCT-US96-00995-3	Sequence 3, Appl
43	94	5.2	1149	1	US-07-915-203-2	Sequence 2, Appl
44	94	5.2	1149	1	US-08-272-887-2	Sequence 2, Appl
45	94	5.2	1149	2	US-08-789-449-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-07-649-591B-3
Sequence 3, Application US/07649591B
Patent No. 5206161
GENERAL INFORMATION:
APPLICANT: Dennis Drayna and Daniel Eaton
TITLE OF INVENTION: NO. 5206161el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/649,591B
FILING DATE: 19910201
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/266-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
US-07-649-591B-3

Query Match 94.5%; Score 1699.5; DB 1; Length 423;
Best Local Similarity 89.8%; Pred. No. 5.5e-175;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 FOSGQVLAALPRISROYQVLT...TTTEIYIMQVTDLTLYKKQVHFFVNASVDVNYKA 60
DB 23 FOSGQVLAALPRISROYQVLT...TTTEIYIMQVTDLTLYKKQVHFFVNASVDVNYKA 82
```

QY 61 HLNVSGLPCSVLLADVEDLIQQOISNDTVSPRASASYEQYHSLNEITYSWIEFTTERHPD 120
DB 83 HLNVSGLPCSVLLADVEDLIQQOISNDTVSPRASASYEQYHSLNEITYSWIEFTTERHPD 142
QY 121 MLTKIHGSSFEKPYLYLVAKSGKEQTAKNAIWDGCIHAREWISPAFCIMFIGH----- 175
DB 143 MLTKIHGSSFEKPYLYLVAKSGKEQTAKNAIWDGCIHAREWISPAFCIMFIGHITQFY 202
QY 176 -----NRMWRKRSFYANNHCIGTDLNRNPAK 203
DB 203 GIIQGYTNLRLVDYVYVMPVYVNDGYDYSWKNNRKRNSFYANNHCIGTDLNRNPAK 262
QY 204 HMCBEGASSSSCSEFTYCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIYFPYSY 263
DB 263 HMCBEGASSSSCSEFTYCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIYFPYSY 322
QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRTYTHGSEFTLYLAPGGDDWITYDGIKYSF 323
DB 323 TRSKSDHEELSLVASEAVRAIEKTSKNTRTYTHGSEFTLYLAPGGDDWITYDGIKYSF 382
QY 324 T 324
DB 383 T 383

RESULT 2
US-08-277-540-3
Sequence 3, Application US/08277540
Patent No. 5474901
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/925-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-277-540-3

Query Match 94.5%; Score 1699.5; DB 1; Length 423;
Best Local Similarity 89.8%; Pred. No. 5.5e-175;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1.

QY 1 FQSGQYLAALPRTSROVOYLONITTYETELVMOPTADLIYKKQVHFFVNASDVDNKA 60
DB 23 FQSGQYLAALPRTSROVOYLONITTYETELVMOPTADLIYKKQVHFFVNASDVDNKA 82
QY 61 HLNVSGLPCSVLLADVEDLIQQOISNDTVSPRASASYEQYHSLNEITYSWIEFTTERHPD 120
DB 83 HLNVSGLPCSVLLADVEDLIQQOISNDTVSPRASASYEQYHSLNEITYSWIEFTTERHPD 142
QY 121 MLTKIHGSSFEKPYLYLVAKSGKEQTAKNAIWDGCIHAREWISPAFCIMFIGH----- 175
DB 143 MLTKIHGSSFEKPYLYLVAKSGKEQTAKNAIWDGCIHAREWISPAFCIMFIGHITQFY 202
QY 176 -----NRMWRKRSFYANNHCIGTDLNRNPAK 203
DB 203 GIIQGYTNLRLVDYVYVMPVYVNDGYDYSWKNNRKRNSFYANNHCIGTDLNRNPAK 262
QY 204 HMCBEGASSSSCSEFTYCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIYFPYSY 263
DB 263 HMCBEGASSSSCSEFTYCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIYFPYSY 322
QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRTYTHGSEFTLYLAPGGDDWITYDGIKYSF 323
DB 323 TRSKSDHEELSLVASEAVRAIEKTSKNTRTYTHGSEFTLYLAPGGDDWITYDGIKYSF 382
QY 324 T 324
DB 383 T 383

RESULT 3
US-08-430-787A-3
Sequence 3, Application US/08430787A
Patent No. 5593674
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616

REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-430-787A-3

Query Match 94.5%; Score 1699.5; DB 1; Length 423;
Best Local Similarity 89.8%; Pred. No. 5.5e-175;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

OY 1 FOSGQVLAALPRTSRQVQLQNTTYYEIVLMQPTADLIYKKQVHFVNASDVNVKA 60
DB 23 FOSGQVLAALPRTSRQVQLQNTTYYEIVLMQPTADLIYKKQVHFVNASDVNVKA 82

OY 61 HLNVSIGPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNIEYSWIEFTTERHPD 120
DB 83 HLNVSIGPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNIEYSWIEFTTERHPD 142

OY 121 MLTKIHSSFEKYPYLVLKVSKEQTAKNAIWDGSHAREWISPAFCIMFIGH----- 175
DB 143 MLTKIHSSFEKYPYLVLKVSKEQTAKNAIWDGSHAREWISPAFCIMFIGHITQFY 202

OY 176 -----NRMNRKRSFYANNHCIGTDLNRNFASK 203
DB 203 GIIGQYTNLRLVDFYVMPVNVNDGYDYSWKKNRMNRKRSFYANNHCIGTDLNRNFASK 262

OY 204 HMCEEGASSSSCSEYTCGLPSEPEVKAVASFLRRINQIKAYISMHSYQHIIVPEYSY 263
DB 263 HMCEEGASSSSCSEYTCGLPSEPEVKAVASFLRRINQIKAYISMHSYQHIIVPEYSY 322

OY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGGDDMIYDLGIKYSF 323
DB 323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGGDDMIYDLGIKYSF 382

OY 324 T 324
DB 383 T 383

RESULT 4
US-08-869-057-2
Sequence 2, Application US/08869057
Patent No. 5983562
GENERAL INFORMATION:
APPLICANT: Morser, Michael J
APPLICANT: Nagashima, Mariko
TITLE OF INVENTION: Method of Detecting Thrombotic Disease
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Berlex Biosciences Legal Department
STREET: 15049 San Pablo Avenue
CITY: Richmond
STATE: California
COUNTRY: USA
ZIP: 94804-0099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,057
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Washtien, Wendy L
REGISTRATION NUMBER: 36,301
REFERENCE/DOCKET NUMBER: 51509AUSM1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-262-5411
TELEFAX: 510-262-7095
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE: Plasma
TISSUE TYPE:
FEATURE:
NAME/KEY: Peptide
LOCATION: 23.401
US-08-869-057-2

Query Match 94.5%; Score 1699.5; DB 2; Length 423;
Best Local Similarity 89.8%; Pred. No. 5.5e-175;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

OY 1 FOSGQVLAALPRTSRQVQLQNTTYYEIVLMQPTADLIYKKQVHFVNASDVNVKA 60
DB 23 FOSGQVLAALPRTSRQVQLQNTTYYEIVLMQPTADLIYKKQVHFVNASDVNVKA 82

OY 61 HLNVSIGPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNIEYSWIEFTTERHPD 120
DB 83 HLNVSIGPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNIEYSWIEFTTERHPD 142

OY 121 MLTKIHSSFEKYPYLVLKVSKEQTAKNAIWDGSHAREWISPAFCIMFIGH----- 175
DB 143 MLTKIHSSFEKYPYLVLKVSKEQTAKNAIWDGSHAREWISPAFCIMFIGHITQFY 202

OY 176 -----NRMNRKRSFYANNHCIGTDLNRNFASK 203
DB 203 GIIGQYTNLRLVDFYVMPVNVNDGYDYSWKKNRMNRKRSFYANNHCIGTDLNRNFASK 262

OY 204 HMCEEGASSSSCSEYTCGLPSEPEVKAVASFLRRINQIKAYISMHSYQHIIVPEYSY 263
DB 263 HMCEEGASSSSCSEYTCGLPSEPEVKAVASFLRRINQIKAYISMHSYQHIIVPEYSY 322

OY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGGDDMIYDLGIKYSF 323
DB 323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGGDDMIYDLGIKYSF 382

OY 324 T 324
DB 383 T 383

RESULT 5
US-08-696-139-2
Sequence 2, Application US/08696139
Patent No. 5672066
GENERAL INFORMATION:
APPLICANT: Fayerman, Jeffrey T.
APPLICANT: Greenen, David P.
APPLICANT: Hershtberger, Charles L.
APPLICANT: Larson, Jeffrey L.
APPLICANT: Steiner, Jane L.
APPLICANT: Zhang, Haichao
TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana

```

RESULT      6
US-08-860-882A-57
; Sequence 57, Application US/08860882A
; Patent No. 5985281
;
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIUS JOHANNES
; APPLICANT: TARACONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS

```

```

Query Match Summary      35.5%; Score 637.5; DB 2; Length 415;
Best Local Similarity    36.0%; Pred. No. 2,6e-60;
Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps
6;

QY      1 FDSGVYLAALPRTSRQVYLAQMLTTTTEYLVMQPVYADLIYVKKQVHFVFNASDVNDYKA 60
Db      20 FEGEKFEYRVNDEHNINIIRELASTPTQIDPFKKPSVSQIQIKHSHYDFEYKAEADYVYEN 79
QY      61 HLNVSQIPCSYLLADVEDDIQQOISNDIVSPASASIVYQOYSLNEIYSWIEFITERHPD 120
Db      80 VKNQELQYKVLISNLRNVEAOFPSRV---RATGSHSEYKRYKMETIEMQTQVATENDPA 136
QY      121 MLTKIHISGFEPKYLPLYLVKSGBDQTKAIAWIDQIGHARWISPAQCIAWTI----- 173
Db      137 LISRVIQIGTFEGRAIYILKV-GKKGQNKPALEIMQCGFHARWISPAQCQWREAVRY 195
QY      174 -----GH-----NRMRKRSFEYANNHCIGTFLNENFASK 203
Db      196 GREIQTVELDKLDEYVLPVLINDGYITWTKSRMRKTRSHIHSSSGICGIDPNRNF-DA 254
QY      204 HMCERGASSSSGSEFYCCGLYPESEFEYVAASFLRLRNINQIKAYTSMHSYQOHIFPYSY 263
Db      255 GWCCEGASRNPCDEYFGCPAAESESETEALADFIHNKLSIKAYLTLSYSQOMMITYPSY 314
QY      264 TRSKSDHEELISVASEVRAIRIEKTSKWTRTYHGHGSETLAPRGCGDMYIDGIKISF 323
Db      315 AYKIGENNAELNALAKAIYKEL-ASLHQTKITTYGCAITTYIPAAGGSDDWAYDQIRYSF 373
QY      324 T 324
Db      374 T 374

```

```

RESULT 7
US-07-649-5918-7
? Sequence 7, Application US/076495918
? Patent No. 5206161
? GENERAL INFORMATION:
? APPLICANT: Dennis Drayna and Daniel Eaton
? TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genentech, Inc.
? STREET: 460 Point San Bruno Blvd
? CITY: South San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94080
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 5.25 inch, 360 KB floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: palin (Genentech)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/07/649, 5918
? FILING DATE: 19910201
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Hasak, Janet E.
? REGISTRATION NUMBER: 28, 616
? REFERENCE/DOCKET NUMBER: 669
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415/266-1896
? TELEFAX: 415/952-9881
? TELEK: 910/371-7168
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 417 amino acids
? TYPE: AMINO ACID
? TOPOLOGY: linear
? US-07-649-5918-7

Query Match 35.2%, Score 632; DB 1; Length 417;
Best Local Similarity 38.2%; Pred. No. 1e-59;
Matches 138; Conservative 57; Mismatches 122; Indels 44; Gaps 7,

QY 1 FQSGGYLAALPRTSQVOYVONLTFTTYEILVMQPTADLIYKKOYHFFVYNASDVNDVKA 60
Db 21 FDREVFVFKVLQNEHNASVKNLTQSIELDFWYPAIHDIANMTYDFVRSKEKESQTIDS 80
QY 61 HUNVSGICSVLADVEDDILQOIS-NDTPSPRASASYEYHSLSNEYISWIEFTTERHP 119
Db 81 TLEQKRIIYEIILHLOEIEIKQFYKDEIAGRHS---YAKINDMDKIVSWTEKMLEKHP 137
QY 120 DMLTRKIHIGSSFEKYPILYLVKYSGEQYAKNAIWDICGIHAREWISPAFLCFI----- 173
Db 138 EMVSRIRIKISFYEDNPILYVTKI-GKKDGBRAKIPMDCCIGHAREWISPARCQFVYQATKS 196
QY 174 -GHN-----RQMKRNRSEFYANNHCIGIDLNRFAS 202
Db 197 YGKNIMTKLLDRMFVYLVPEFNVGDYIWSWTQDRMRKRNRSRNSNSTCIGIDLNNF-D 255
QY 203 KHWCEGASASSSCSEYCYGLVPESEPEYKAVASFLRRNINOIKAYTISMHSYSOHIYFVS 262
Db 256 VSMDSPPMTNPKCLANVYIGPAPASEKEKRYAVNTFIRSHLNSIKAYITTFISYSOMLPIYG 315
QY 263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDMWTYDGIKYS 322
Db 316 YTFKLPNNHQLLKVARIATDAL-STREYETRYLYGPIASTIKYTGSSSLDWVYDGIKHT 374
QY 323 F 323
I

```

```

Db      375 F 375

RESULT      8
US-08-277-540-7
: Sequence 7, Application US/08277540
: Patent No. 5474901
: GENERAL INFORMATION:
:   APPLICANT: Drayna, Dennis T., Eaton, Dan L.
:   TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
:   NUMBER OF SEQUENCES: 8
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Genentech, Inc.
:     STREET: 460 Point San Bruno Blvd
:     CITY: South San Francisco
:     STATE: California
:     COUNTRY: USA
:     ZIP: 94080
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: 5.25 Inch, 360 Kb Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: patin (Genentech)
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/277,540
:     FILING DATE: 19-JUL-1994
:     CLASSIFICATION: 435
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: 08/167727
:       FILING DATE: 15-DEC-1993
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: 07/959944
:       FILING DATE: 14-OCT-1992
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: 07/649591
:       FILING DATE: 01-FEB-91
:     ATTORNEY/AGENT INFORMATION:
:       NAME: Hasak, Janet E.
:       REGISTRATION NUMBER: 28,616
:       REFERENCE/DOCKET NUMBER: 689DJC1D1
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: 415/225-1896
:       TELEFAX: 415/952-9881
:       TELEX: 910/371-7168
:     INFORMATION FOR SEQ ID NO: 7:
:       SEQUENCE CHARACTERISTICS:
:         LENGTH: 417 amino acids
:         TYPE: amino acid
:         TOPOLOGY: linear
:       US-08-277-540-7

Query Match      35.2%, Score 632; DB 1; Length 417;
Best Local Similarity 38.2%, Pred. No. 1e-59;
Matches 138; Conservative 57; Mismatches 122; Indels 44; Gaps 7;

OY      1 FOSGGVLAALPPTSQOVOLNLTITTYETVLMQPTADLIYKKOVHFFVNASDVDNKA 60
      21 FDREVFVRYKQLONEKHAASYLKLNTGSIELDFWYPAIHDIANMTVDPRVSKESQTTQS 80
      61 HLNVSIGPCSVLLADVEDLLIQOQIS-NDTVSPRASASYEEOYHSINLETIYSWIEFTTERHP 119
      81 TLEQKRIHYEILLHDLEIEIKQFVDVDEIAGRHS--YAKYNDMDKIIVSMEKMLEKHP 137
OY      120 DMLTPIHIGSGSEKPLIYLVKVSAGEQJAKNAIWIIDCGIHAREWISPAFCIMFT----- 173
      138 EMVSVIRIKIGISTVEDNPVLVAKI-GKKDERKAIIFMDCGIHAREWISPAFCQWFFYQATRS 196
OY      174 -GHN-----RWRRKNSFEYANNNGICGTDLNRNFS 202
      197 YGKNKIMTKLDRNMFYVLPVPFNVGDIYTSWTQDMMKKNRNRNNSICIGTDLNRNF-D 255
OY      203 KHWCEGASSSSCSETYGLYPESEPEYKAVASFLRRRINIQIKAVISMHSYSQHIIVFPYS 262

```



```

0Y      61  HUNSGJPCSVLADVEEDL00Q01S-DJWSPRASASYEQJSHSLNFIEMETERHP 119
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db      81  ALDONKMHYTELLIHDLOPEIKOPDYVEDIDPGRHS---YAKINMEYIYAMTERKMDKP 137

0Y      120  DMLTRKHIGHSFEKYPPLYVLKVSCKEQTAKNAIIMDGIHAREWISPAFLAMEY----- 173
      | : : : : | | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      138  EMWSIRKIGSVIEDNPPLYLTKI-GEKNRRRAIRIMDCGIHAREVSPAFQWYQATKT 196

0Y      174  -----GH-----NRMRKNRSFYANNHCIGTDLNRNFPAS 2020
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db      197  YGRNKKIMTKLIDRNNFYTLIPVFNVDGYIWSWTKRMRMRKNRSKQNSKJCIGTDLNRNF-N 255

0Y      203  KHNCEGASSSSCSEFYCGLPESEPEYKAVASFLRNNINOIKAYISMHSYSGHIVPYS 262
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db      256  ASWMSIPTNPDPCADNRYGASPESEKTEKTAVTNPIRSHLEIKYIIFHSYSOMLLEPYG 315

0Y      263  YTRRSKSDHEHLSIVASPAVAIKETSKNRYRTHGHSFLLYLAPGGGDWIMYIDLGIKYS 322
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db      316  YTSKLPRNHEDLAVAKIGTDVLT-STREYKTYIYGPIESTIYPIGSSILDMAYDLGIKHT 374

0Y      323  F 323
      |
Db      375  F 375

```

RESULT 11
US-08-277-540-6
: Sequence 6, Application US/08277540
: Patent No. 5474901
: GENERAL INFORMATION:
: APPLICANT: Drayna, Dennis T., Eaton, Dan L.
: TITLE OF INVENTION: NO. 5474901el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/277, 540
: FILING DATE: 19-JUL-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/167727
: FILING DATE: 15-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/959944
: FILING DATE: 14-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/649591
: FILING DATE: 01-FEB-91
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28, 616
: REFERENCE/DOCKET NUMBER: 68991C1D1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 417 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
US-08-277-540-6

Query Match	34.68;	Score 622;	DB 1;	Length 417;
Best Local Similarity	35.58;	Pred. No. 1.2e-58;		
Matches 128;	Conservative 71;	Mismatches 118;	Indels 44;	Gaps 7

OY		1	FOSQOVLALPRTSROVOVLONLTFTTEIYLMDCPVADLLIVKKQVAFFPNASVDNVKA	60
Dd		21	FDRREKVEYRVKPQDEKQADI IKDLAKTIDELDEWYGATIHHAANMVDFRSEKESQAIGS	80
OY		61	HLNVSIGPCSVLADVDELIOOOIS -NDVSPRASASYEQYHSLNEIYSWIEETERHP	119
Dd		81	ALDONKMHHYEILLHDLOEEIEKEFDVKEDIPGRHS--YAKYNMEKIYAMTEKMDMKYP	137
OY		120	DMLTRKHGSGFEKYPTLYLVKSGEKQTAKNAITDGGIARENISAFCLTFPI-----	173
Dd		138	EMVASRIKIGSTVEDNPPLYVLI-KGNERRKRAIFMDGIIHAREVWSAFCOMFVYOATKT	196
OY		174	-----GH-----NRMRKRNSFYANHRCIGTDLNNRFAS	202
Dd		197	YGRNKIMTKLLDRMNPFYILPFPNVDCGIWSTKTRRMKRNRKSQNOKSCIGTDLNNRF-N	255
OY		203	KHMCEGASSSSCSETYCGLYPESEPZYKAVASFELRRNIQIKAYISMHSYOHIEPYPS	262
Dd		256	ASWMSIPTNDPCADNYRGSAPESEKETKCAVTNPIRSHLNIEIKYILFIHFSYQMLLEPYG	315
OY		263	YTRSKSDHEHLSLVASEAVALIKTSKNTRYTHGHGSSEFLYAPAGGGDDMIYDLGIKIS	322
Dd		316	YTSKLPRNHEDLAKAVAKIGDIVL-STRETYEYYIGPIESTIYPIPGSSSLDMAYDLGIKHT	374
OY		323	F	323
Dd		375	F	375

US-08-430-787A-6
Sequence 6, Application US/08430787A
Patent No. 5593674
GENERAL INFORMATION:
APPLICANT: Deyana, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/6649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616

```

1      RESULT 13
2      US-09-171-945-125
3      : Sequence 125 Application US/09171945
4      : Patent No. 6277599
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Emery, Stephen
9      : APPLICANT: Copley, Clive Graham
10     : APPLICANT: Edge, Michael Derek
11     :
12     : TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said
13     : FILE REFERENCE: Antibody, and Their Therapeutic Use in an Adept System
14     : CURRENT APPLICATION NUMBER: US/09/171,945
15     : CURRENT FILING DATE: 1998-10-29
16     : PRIOR APPLICATION NUMBER: GB9703103.3
17     : PRIOR FILING DATE: 1997-02-14
18     : PRIOR APPLICATION NUMBER: GB9609405.7
19     : PRIOR FILING DATE: 1996-05-04
20     : PRIOR APPLICATION NUMBER: PCT/GB97/01165
21     : PRIOR FILING DATE: 1997-04-29
22     : NUMBER OF SEQ ID NOS: 131
23     : SOFTWARE: PatentIn Ver. 2.1
24     : SEQ ID NO 125
25     : LENGTH: 716
26     : TYPE: PRT
27     : ORGANISM: Artificial Sequence
28     : FEATURE:
29     : OTHER INFORMATION: Description of Artificial Sequence: humanized

```

```

QY      1 FOSGOVLAALPRTSKROVOVLONLITTYEYLIVMOPTADILAKKQOVHFFVNASDVNYKA    60
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      22 FEGEKEFRVNVEDENHNINIRELASTQTIDFCKPRDSVTQLIRPHSTVDPRVAEDTUYVEN    81
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      61 HUNVSQIPCVSLAVAYVEDLLOOQIINDPVSPRASASTYEOTHLSNEIYSWIEFTTERHPD    120
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      82 YAKOMELOKYVLISLRNVEAEQFDSDRV--RATGHSETEKNMKETTEAMTOQVATENPA    138
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121 MLTKIHIGSSPEKRYLYLVKASGKEQOTAKNAIMWIDCGFHAREWISPPCIMEFI-----    173
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      139 IHSRSVIGTFEGRAIYILKV-CRAGCOKKPALIPMDCGFHAREWISPPACQFNEAVRTY    197
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      174 -----GH-----NNRMKRNSFYANNICIGTDILNRFPASK    203
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      198 GREIQVELLDKLDPYVLPVLINDIGYITWTXSRFMKTRSTHTHGSSCIGTPDRNR-DA    256
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      204 HMCEBGASSSCSETYGCGLYPESEDEPKAVASFLRRINIOLIKATISMHSTSQHUTVPYST    263
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      257 GWCGLGABRNMCDELEYCPAAESKEKETKALADFLRNLISIKRAYLITHSOSQMUIPYSV    316
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      264 TRSKSKDHIELSLVASAEVRAIERKTSKNTRTTHGSHGETSYLAPGGGDMDIYDLGIKSPS    323
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      317 AYKLGENNALNALAKATVKEL-ASLHGTVYTGPGRATITYPSAGNSTKDMAIYDOGITRSE    375
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      324 T 324
Db      |
QY      376 T 376

```

RESULT 14
 US-08-782-760-6
 : Sequence 6, Application US/08782760
 : Patent No. 5948668
 :
 : GENERAL INFORMATION:
 : APPLICANT: Hartman, Jacob
 : APPLICANT: Fulga, Netta
 : APPLICANT: Mendelovitch, Simona
 : APPLICANT: Gorecki, Marian
 : TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
 : TITLE OF INVENTION: CARBOXYPEPTIDASE B
 : NUMBER OF SEQUENCES: 8
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Cooper & Dunham LLP
 : STREET: 1185 Avenue of the Americas
 : CITY: New York
 : STATE: New York
 : COUNTRY: U.S.A.
 : ZIP: 10036
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/782,760
 : FILING DATE: 13-JAN-1997
 : CLASSIFICATION: 435
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/378,233
 : FILING DATE: 25-JAN-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: White, John P.
 : REGISTRATION NUMBER: 28,678
 : REFERENCE/DOCKET NUMBER: 0336/43847
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (212) 278-0400

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:45:11 ; Search time 65 Seconds

(without alignments)
577.584 Million cell updates/sec

Title: us-09-980-881-3

Perfect score: 1798
Sequence: 1 FQSGQVLAALPRISROYVLT.....IKTSFTSNPEVKLLPLSLK 338

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008
Listing first 45 summaries

Database :

A.Geneseq_032802:*
1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1798	100.0	338	21	AAAB1458 Human brain carbox
2	1785	99.3	360	21	AAAB1457 Human brain carbox
3	1699.5	94.5	423	17	AAAR90293 Human plasma carbo
4	1699.5	94.5	423	18	AAAW14733 Human plasma carbo
5	1699.5	94.5	423	20	AAW92270 Human plasma carbo
6	1694.5	94.2	423	14	AAAR36273 Human plasma carbo
7	1338	74.4	246	21	AAAB1459 Human brain carbox
8	816.5	45.4	211	21	AAAB58129 Lung cancer associ
9	646.5	36.0	404	16	AAAR75131 Porcine Tyr-His-Me
10	637.5	35.5	415	17	AAW06172 Human pancreatic c
11	637.5	35.5	415	19	AAW74476 Human pancreatic c

12	637.5	35.5	424	17	AAW06175	ProHCPB with PeLB
13	636.5	35.4	324	22	AAAG65562	Bothops jataraca c
14	619.5	34.5	716	18	AAW41414	PreproHCPB-linker-
15	612	34.0	437	22	AAE00502	Human carboxypepti
16	610	33.9	307	17	AAW00602	Rat carboxypeptida
17	582	32.4	306	16	AAW75132	Porcine carboxypep
18	582	32.4	306	22	AAW04477	Human secreted met
19	569	31.6	354	22	AAAG65560	Human secreted met
20	569	31.6	354	22	AAAG65566	Human secreted met
21	569	31.6	354	22	AAAG65566	Human secreted met
22	569	31.6	354	22	AAAG65567	Human secreted met
23	569	31.6	354	22	AAAG65567	Human secreted met
24	567	31.5	329	17	AAW06174	Mature HCPB with P
25	567	31.5	349	17	AAW06173	Modified HCPB (H1s)6
26	560	31.1	349	17	AAW06181	Modified HCPB (D25)
27	560	31.1	349	18	AAW13749	Modified HCPB (D25)
28	559	31.1	349	17	AAW06182	Modified HCPB (D25)
29	559	31.1	349	18	AAW13750	Carboxypeptidase B
30	556	30.9	349	18	AAW13750	Carboxypeptidase B
31	554	30.8	349	18	AAW13753	Carboxypeptidase B
32	554	30.8	349	18	AAW13752	Carboxypeptidase B
33	554	30.8	349	18	AAW13752	Carboxypeptidase B
34	553	30.8	349	18	AAE01664	Novel human protea
35	552	30.7	349	18	AAW13753	Carboxypeptidase B
36	551	30.6	311	22	AAAG65563	Carboxypeptidase B
37	551	30.6	349	18	AAW13760	Human carboxypepti
38	551	30.6	349	18	AAW13761	Carboxypeptidase B
39	551	30.6	349	18	AAW13762	Carboxypeptidase B
40	549	30.5	349	18	AAW13757	Carboxypeptidase B
41	549	30.5	349	18	AAW13765	Carboxypeptidase B
42	549	30.5	613	19	AAW41412	Humanised Fd-mutan
43	548	30.5	613	19	AAW74481	Human pancreatic c
44	548	30.5	349	18	AAW13759	Carboxypeptidase B
45	546	30.4	349	18	AAW13764	Carboxypeptidase B

ALIGNMENTS

RESULT 1	AAAB1458	standard: Protein; 338 AA.
ID	AAAB1458	
AC	AAAB1458	
XX		
DT	01-MAR-2001	(first entry)
XX		
DE	Human brain carboxypeptidase B protein SEQ ID NO 3.	
XX		
KW	Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;	
KW	treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;	
KW	cerebroprotective; antialzheimers; nootropic; neuroprotective;	
KW	hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;	
KW	Down's syndrome; head trauma.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200066717-A1.	
XX		
PD	09-NOV-2000.	
XX		
PF	01-MAY-2000; 2000WO-JF02878.	
XX		
PR	30-APR-1999; 99JP-0125169.	
XX		
PA	(MATS/) MATSUMOTO A.	
XX		
PI	Matsumoto A;	
XX		
DR	WPI; 2000-687534/67.	
XX		
PT	Human brain carboxypeptidase B isolated from the hippocampus useful for screening agents for the treatment of Alzheimer's and other brain	

PT disorders -
 XX
 PS Disclosure; Page 71-73; 84pp; Japanese.
 CC
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 CC
 XX
 SO Sequence 338 AA;

Query Match 100.0%; Score 1798; DB 21; Length 338;
 Best Local Similarity 100.0%; Pred. No. 1.2e-179;
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQSGVLAALPRTSRQVYQVQNTTTEYIVLMQPTADLVKKQVHFVNASDVNVKA 60
 DB 1 fgsqgvlaalprrtsrqvqvlqnltyeyvlwqvtcadllvkkqvhffvnasdvndvka 60
 QY 61 HLNVSGIPGSVLLADVEDLIQOOISNDTVSPRASASYEYQHSINETYMTETTERHPD 120
 DB 61 hlrvsglpsavlladvedllqgqisndtvsprrasasyeqhnslnetywlefterhpd 120
 QY 121 MLTKIHIGSSFEKYPVLYLVKVGSKQETAKNAIWDGCIHAREWISPAFLMFTIGHNRMWR 180
 DB 121 mltkihlgssfekyplylvkvgskqetaknaikwdcgiharewispafclwflghnmwr 180
 QY 181 KNRSEYANNHCIGTDLRNPFASKHMCCEGASSSCSETYCGILPESPEPVKAVASFLRRN 240
 DB 181 knrsfyanhncigtldlrnrfaskhmcceegassscsetyqgilypsepepvkavasflrrn 240
 QY 241 INQIKAYISMHSYQHIIVFPYSYTRSKSKDHEELSLVAEAVRAIEKTSKNTRYTHGHS 300
 DB 241 inqikayismhsyqhiivfpystrskskdheelslvaseavraidektskntrythghs 300
 QY 301 ETLYLAPGGGDWYIDLGIKYSFTSNPPVEKLLPLSLK 338
 DB 301 etlylapgggdwyidlgikysftsnppveklplslk 338

RESULT 2
 AAB1457
 ID AAB1457 standard; Protein: 360 AA.
 AC AAB1457;
 XX
 DT 01-MAR-2001 (first entry)
 XX
 DE Human brain carboxypeptidase B protein.
 XX
 KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KW cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
 KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KW Down's syndrome; head trauma.
 XX
 OS Homo sapiens.
 XX
 PN W0200066717-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 01-MAY-2000; 2000MO-JP02878.
 XX
 PR 30-APR-1999; 99JP-0125169.
 XX
 PA (MATS/) MATSUMOTO A.

XX
 PI Matsumoto A;
 XX
 DR WPI: 2000-687534/67.
 DR N-PSDB: AAC81962.
 XX
 PT Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 XX
 PS Claim 1: Page 68-71; 84pp; Japanese.
 XX
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 CC
 XX
 SO Sequence 360 AA;

Query Match 99.3%; Score 1785; DB 21; Length 360;
 Best Local Similarity 99.1%; Pred. No. 3e-178;
 Matches 335; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FQSGVLAALPRTSRQVYQVQNTTTEYIVLMQPTADLVKKQVHFVNASDVNVKA 60
 DB 23 fgsqgvlaalprrtsrqvqvlqnltyeyvlwqvtcadllvkkqvhffvnasdvndvka 82
 QY 61 HLNVSGIPGSVLLADVEDLIQOOISNDTVSPRASASYEYQHSINETYMTETTERHPD 120
 DB 83 hlrvsglpsavlladvedllqgqisndtvsprrasasyeqhnslnetywlefterhpd 142
 QY 121 MLTKIHIGSSFEKYPVLYLVKVGSKQETAKNAIWDGCIHAREWISPAFLMFTIGHNRMWR 180
 DB 143 mltkihlgssfekyplylvkvgskqetaknaikwdcgiharewispafclwflghnmwr 202
 QY 181 KNRSEYANNHCIGTDLRNPFASKHMCCEGASSSCSETYCGILPESPEPVKAVASFLRRN 240
 DB 203 knrsfyanhncigtldlrnrfaskhmcceegassscsetyqgilypsepepvkavasflrrn 262
 QY 241 INQIKAYISMHSYQHIIVFPYSYTRSKSKDHEELSLVAEAVRAIEKTSKNTRYTHGHS 300
 DB 263 inqikayismhsyqhiivfpystrskskdheelslvaseavraidektskntrythghs 322
 QY 301 ETLYLAPGGGDWYIDLGIKYSFTSNPPVEKLLPLSLK 338
 DB 323 etlylapgggdwyidlgikysftsnppveklplslk 360

RESULT 3
 AAR90293
 ID AAR90293 standard; Protein: 423 AA.
 AC AAR90293;
 XX
 DT 12-APR-1996 (first entry)
 XX
 DE Human plasma carboxypeptidase B.
 XX
 KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
 KW purification; plasminogen; affinity column.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Active-site 181
 FT /note= "Residue suspecting of being involved in
 FT catalytic activity of enzyme."

FT	Binding-site	183	/note-	"Residue suspected of being part of the binding site of enzyme."
FT	Binding-site	256..257	/note-	"Residues suspected of being part of the binding site of enzyme."
FT	Active-site	310	/note-	"Residue suspected of being involved in catalytic activity of enzyme."
FT	Binding-site	312	/note-	"Residue suspected of being part of the binding site of enzyme."
FT	Active-site	369	/note-	"Residue suspected of being involved in catalytic activity of enzyme."
FT	Binding-site	370..371	/note-	"Residues suspected of being part of the binding site of enzyme."
FT	Binding-site	394	/note-	"Residue suspected of being part of the binding site of enzyme."
FT	US5474901-A.			
PD	12-DEC-1995.			
PE	01-FEB-1991;	91US-0649591.		
PR	01-FEB-1991;	91US-0649591.		
PR	14-OCT-1992;	92US-0959944.		
PR	15-DEC-1993;	93US-0167727.		
PR	19-JUL-1994;	94US-0277540.		
PA	(GETH) GENENTECH INC.			
PI	Drayna DT, Eaton DL;			
DR	WPI; 1996-039508/04.			
DR	N-PSDB; AAT11671.			
XX	Antibody to human plasma carboxypeptidase B - useful for detecting and purifying hPCPB for use in treating clotting disorders e.g. haemophilia A			
XX	Disclosure; Figure 4; 40pp; English.			
XX	An antibody which specifically binds human plasma carboxypeptidase B (hPCPB) and does not cross react with other carboxypeptidases is useful for the detection of hPCPB in vitro. The antibody is also used for purifying hPCPB from a sample. Purification comprises passing a sample thought to contain hPCPB over either a column to which antibody has been bound, or a plasmidogen affinity column, eluting the column and then recovering the fraction containing the hPCPB.			
XX	Sequence	423 AA;		
XX	Query Match	94.15%;	Score 1699.5;	DB 17; Length 423;
XX	Best Local Similarity	89.88%;	Pred. No. 3.6e-169;	
XX	Matches 324; Conservative	0;	Mismatches 0;	Indels 37; Gaps 11.
QY	1 FOSGGVLAALRRTSRKOVVLONLTTTTEVLMOPTADLIVRKKOVHFFVNASDVNYKA	60		
DB	23 fsgsgvialalprtsrqvqlnlttlyelivkqvltadiivkkkyhffvnasdvnyka	82		
OY	61 HLNAGSICPSVILADVEDLIQQOISNDTVSPRASAYYEQYHSLNEIYSWIEFTERRHPD	120		
DB	83 hlnvsglpcsvlladvedliqqisndtvspraasayeyqylsneiyswiefierhpd	142		
OY	121 MLTKHIGSSFEKRYLYLVLYKXSGKQOTAKNAIWDICGHARWRISPAFLWFIQH-----	175		
DB	143 mltklhgssfekrpylylvlykxsgkqaknaIwdicgharwispafclwfiqhtqfy	202		

Oy	176	-----NMAMKNSPFAANNICITDINRNPAK	203
Dd	203	gIlgqYcnllIlvdfymrvvvndgdyswknmmxkmsIfyanmctglalnrfask	262
Oy	204	HMCESGSSSCSEFTYGLYPESEPEVKAVASPLRRNIQIKAYISMHSYSQHIVEPYR	263
Dd	263	hwceagassscsetylipsepevkavasflrrnlnqIkayImhsysqlvlfpys	322
Oy	264	TRRSKDHEELSLVASEAVALIENTSKNTRYTHGHSETLYLAPGGDDMIYDLGIKRS	323
Dd	323	trrskdheelslvaseavraliektskntrytghngsetlylapggddviylgiikystf	382
Oy	324	T 324	
Dd	383	t 383	
 RESULT 4 AAW14733			
ID	AAW14733	standard; Protein; 423 AA.	
XX	AAW14733;		
AC			
XX			
DT	08-MAY-1997	(first entry)	
XX			
DE	Human plasma carboxypeptidase B.		
XX			
KW	Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;		
KM	plasma; plasminogen.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1..22	
FT		/note- "Signal peptide"	
FT	Protein	23..423	
FT		/note- "Mature PCPB"	
FT	Cleavage-site	114	
FT		/note- "Potential clip site for activation of PCPB as a carboxypeptidase"	
FT	Active-site	181	
FT		/note- "Expected to be involved in catalytic activity"	
FT	Binding-site	183	
FT		/note- "Expected to be involved in substrate binding"	
FT	Active-site	184	
FT		/note- "Expected to be involved in catalytic activity"	
FT	Binding-site	256	
FT		/note- "Expected to be involved in substrate binding"	
FT	Binding-site	257	
FT		/note- "Expected to be involved in substrate binding"	
FT	Active-site	310	
FT		/note- "Expected to be involved in catalytic activity"	
FT	Binding-site	312	
FT		/note- "Expected to be involved in substrate binding"	
FT	Active-site	369	
FT		/note- "Expected to be involved in catalytic activity"	
FT	Binding-site	370	
FT		/note- "Expected to be involved in substrate binding, thought to determine specificity of PCPB as a carboxypeptidase B"	
FT	Binding-site	371	
FT		/note- "Expected to be involved in substrate binding"	
FT	Binding-site	394	
FT		/note- "Expected to be involved in substrate binding"	
XX			
PN	US5593674-A.		
XX			
PD	14-JAN-1997.		
XX			
PR	01-FEB-1991;	91US-0649591.	
XX			
PR	01-FEB-1991;	91US-0649591.	
XX			

PR 14-OCT-1992; 92US-0959944.
 PR 15-DEC-1993; 93US-0167727.
 PR 19-JUL-1994; 94US-0277540.
 PR 27-APR-1995; 95US-0430787.

XX (GETH) GENENTECH INC.

PI Drayna DT, Eaton DL;

XX MPI; 1997-099413/09.

DR N-PSDB; AAT62846.

XX Using human plasma carboxypeptidase B in blood coagulation - is

PT functionally related to carboxypeptidase A and pancreas

XX carboxypeptidase B

XX Example 2; Column 37-42; 39pp; English.

CC This sequence represents human plasma carboxypeptidase B (PCPB) which
 CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
 CC PCPB may be used therapeutically in haemostatic regulation. PCPB is
 CC purified from human plasma or by transformed cell culture by
 CC extraction using plasminogen bound to a solid phase.

XX Sequence 423 AA;

Query Match 94.5%; Score 1699.5; DB 18; Length 423;
 Best Local Similarity 89.8%; Pred. No. 3,6e-169;

Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 FOSGOVLAALPRTSRQVOVLQNTTTEYELVMOPTADLIYKKQYHFFVNASDVNDYKA 60
 DB 23 fgsqgylaalprrtsrqvqlnlttelyelwqvtadliivkkqyhfivnasdvndyka 82
 QY 61 HUNSGIPC SVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNEIYSWIEFITERHPD 120
 DB 83 hlnvsgipcsvlladvedliqgqindtvsprasaasyeqyhslnelyswiefiterhp 142
 QY 121 MLTKIHGSSFEKYPYLVKVGSGEOTAKNAWIDCGIHAREWISPAFLMFTIGH----- 175
 DB 143 mltkihgssfekyplylvkvsqgeaknalwldcgiharewispafclwflghltqfy 202
 QY 176 -----NRMARKNSFYANNHCIGTDLNRNFASK 203
 DB 203 gllgqytnllrlvdfymvrvnvdydyswkknmrknsfyannhcgtdlnrnfask 262
 QY 204 HMCEGASSSSCSEYTCGLYPESEPEVKAVASFLRRNINQIKAYISMSHSOHIYFPYSY 263
 DB 263 hwceegasssscsetcglypesepevkavasflrrninqikayismhsyqhlvfpysy 322
 QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSEFTLVLAGGDDWIVDGIKYSF 323
 DB 323 trsksdheelslvaseavraiektkskntcrythngsecltlapggddwivdgiikysf 382
 QY 324 T 324
 DB 383 t 383

RESULT 5

AAW92270 ID AAW92270 standard; Protein; 423 AA.

XX AAW92270;

XX 28-APR-1999 (first entry)

DE Human plasma carboxypeptidase B (PCPB) chr147.

KW Plasma carboxypeptidase B; PCPB; human; hPCPBchr147;

KM polymorphism detection; thrombotic disease.

XX

OS Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..22

XX Protein /note="signal peptide"

XX 23..423 /note="mature PCPBchr147"

PN W09855645-A1.

XX 10-DEC-1998.

XX 02-JUN-1998; 98MO-EP03244.

XX 03-JUN-1997; 97US-0869057.

XX (SCHD) SCHERING AG.

XX Morser MJ, Nagashima M;

XX MPI; 1999-045800/04.

XX N-PSDB; AAV74302.

PT Detecting new polymorphism of human plasma carboxypeptidase B -
 PT comprises Alanine or Threonine at position 147 of protein by DNA or
 PT protein analysis, useful to detect risk of thrombotic disease in
 PT humans

PS Disclosure: Page 25-26; 35pp; English.

CC This sequence is the human plasma carboxypeptidase B (PCPB) mutant
 CC hPCPBchr147. The invention relates to a method for determining the
 CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
 CC comprises obtaining a prepared tissue or blood sample and determining the
 CC presence of DNA coding for naturally occurring polymorphisms of the protein
 CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
 CC respectively). Determination of the relative distribution of the PCPB
 CC polymorphs in a patient's blood by genetic or protein analysis by the
 CC methods is useful to determine the risk of thrombotic disease in humans.
 CC Such assessments may be made by accumulating information concerning the
 CC relative distribution of the different polymorphs within the general
 CC population compared with populations known to be at risk and establishing
 CC a PCPB polymorph profile for at-risk patients.

XX Sequence 423 AA;

Query Match 94.5%; Score 1699.5; DB 20; Length 423;
 Best Local Similarity 89.8%; Pred. No. 3,6e-169;

Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 FOSGOVLAALPRTSRQVOVLQNTTTEYELVMOPTADLIYKKQYHFFVNASDVNDYKA 60
 DB 23 fgsqgylaalprrtsrqvqlnlttelyelwqvtadliivkkqyhfivnasdvndyka 82
 QY 61 HUNSGIPC SVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNEIYSWIEFITERHPD 120
 DB 83 hlnvsgipcsvlladvedliqgqindtvsprasaasyeqyhslnelyswiefiterhp 142
 QY 121 MLTKIHGSSFEKYPYLVKVGSGEOTAKNAWIDCGIHAREWISPAFLMFTIGH----- 175
 DB 143 mltkihgssfekyplylvkvsqgeaknalwldcgiharewispafclwflghltqfy 202
 QY 176 -----NRMARKNSFYANNHCIGTDLNRNFASK 203
 DB 203 gllgqytnllrlvdfymvrvnvdydyswkknmrknsfyannhcgtdlnrnfask 262
 QY 204 HMCEGASSSSCSEYTCGLYPESEPEVKAVASFLRRNINQIKAYISMSHSOHIYFPYSY 263
 DB 263 hwceegasssscsetcglypesepevkavasflrrninqikayismhsyqhlvfpysy 322
 QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSEFTLVLAGGDDWIVDGIKYSF 323
 DB 323 trsksdheelslvaseavraiektkskntcrythngsecltlapggddwivdgiikysf 382

Dd	323	trskskdhseelsvaseavraiektsnrtlytghnsetlylapggddwiydlgkysf	382
Oy	324	T 324	
Dd	383	t 383	
RESULT 6			
ID	AAR36273	standard; Protein; 423 AA.	
XX	AAR36273;		
AC	AAK36273;		
XX	24-AUG-1993	(first entry)	
DT			
XX	Human plasma carboxypeptidase B.		
DE	PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.		
XX			
KW	Homo sapiens.		
OS			
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1..22	
FT	Cleavage-site	/label= signal-peptide 114	
FT		/note= "putative site of cleavage by trypsin to activate PCPB as a carboxypeptidase"	
FT	Active-site	181	
FT		/note= "expected to be involved in catalytic activity"	
FT	Binding-site	183	
FT		/note= "expected to be involved in substrate binding"	
FT	Active-site	184	
FT		/note= "expected to be involved in catalytic activity"	
FT	Binding-site	256..257	
FT		/note= "expected to be involved in substrate binding"	
FT	Active-site	310	
FT		/note= "expected to be involved in catalytic activity"	
FT	Binding-site	312	
FT		/note= "expected to be involved in substrate binding"	
FT	Active-site	369	
FT		/note= "expected to be involved in catalytic activity"	
FT	Binding-site	370..371	
FT		/note= "expected to be involved in substrate binding; Asp 370 is the residue which determines specificity of the PCPB as a carboxypeptidase B"	
FT	Binding-site	394	
FT		/note= "expected to be involved in substrate binding"	
XX			
PN	US5206161-A.		
XX	27-APR-1993.		
PD			
XX	01-FEB-1991;	9IUS-0649591.	
PF			
PR	01-FEB-1991;	9IUS-0649591.	
XX			
PA	(GETH) GENENTECH INC.		
PI	Drayna Dr., Eaton Dc;		
XX			
DR	WPI: 1993-151724/18.		
XX	N-PSSD; AAQ41001.		
XX	New human plasma carboxypeptidase B - used as haemostatic		

PT		regulator for clotting blood, partic. for treating blood clotting disorders, e.g. haemophilia
XX		
PS	Claim 2; Fig 4; 40pp; English.	
XX		
CC	This amino acid sequence was deduced from a clone isolated from a human liver cDNA library. There is about 40% sequence identity between the deduced preproPCB amino acid sequence and pro-human mast cell carboxypeptidase A and between preproPCB and prepro-rat carboxypeptidase B. Human PCB has the same substrate binding sites as, and shares 6 cysteine residues which form 3 intramolecular disulphide bonds with, bovine and rat carboxypeptidase B. The presence of the same amino acid (Asp at position 348 of the mature protein) at the region in carboxypeptidases that determines substrate specificity, suggests that PCB represents a plasma-derived carboxypeptidase. PCPB inhibits the enzymatic conversion by tPA of plasminogen to plasmin in the presence of fibrinogen.	
CC		
CC		
CC		
CC		
CC		
XX	Sequence 423 AA:	
SQ		
Query Match	94.2%; Score 1694.5; DB 14; Length 423;	
Best Local Similarity	89.5%; Pred. No. 1.2e-168;	
Matches 323; Conservative	0; Mismatches 1; Indels 37; Gaps 1;	
OY	1 FOSGIVLAALPRTSRKOVYLQMLTTTYEILVLMQPVTADLIVKKGVHFFNASDVNNKA 60	
DB	23 fsggylaalprtsrqvqlgnlttelylwkpvrtadivkkkyhffnasdvnnka 82	
OY	61 HUNSGIPCSSLADVEDLIQQOISNDIYSPRASASYEQYHSLEIYSWIEFTERRHPD 120	
DB	83 hlnvgslpcsvlladvedliqgqindelvspasasyeqhnlleylswiefterhp 142	
OY	121 MLTKIHGSEFKPYLYLVKSKGEOTAKNALIWIDCGIHAREWISAPCLMTIRGH---- 175	
DB	143 mltklhigstefkyplylvksqgeqtaknalwldcgiharewisapclwlfightqfy 202	
OY	176 -----NRMMKRNSFYANNHCIGTDLNNNFASK 203	
DB	203 gllgqytullrlvydymrvnvvdgydykwknrmwrkirsfyannhcigtclnfnfask 262	
OY	204 HMCEBGASSCSSESYCYGLYPESEPVKAVASFLLRNINQIKAYISMHSYOHIVPPSY 263	
DB	263 hmcebgacsscsseeycglypesepvkavasflrrnlnqikayismhsyqhivfpysy 322	
OY	264 TFSKSDHEELSLVASEANRALETEKTSKNRTRHGSGETLYLARPGGDWITVDLGIKYSF 323	
DB	323 trfskdheelslvaseanraletekstsknrtrhgsetlylarpggdwytdlgikyaf 382	
OY	324 T 324	
DB	383 t 383	
RESULT 7		
AAB11459		
ID	AAB11459 standard; Protein; 246 AA.	
XX	AAB11459;	
AC		
XX		
DT	01-MAR-2001 (first entry)	
XX		
DE	Human brain carboxypeptidase B protein SEQ ID NO 4.	
XX		
KM	Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase; treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine; neuroprotective; anti-Alzheimer's; nontoxic; neuroprotective;	
KM	hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage; Down's syndrome; head trauma.	
KX		
XX	Homo sapiens.	
OS		
XX	WO200066717-A1.	
PN		

```

XX 09-NOV-2000.
PD
XX
XX 01-MAY-2000; 2000MO-JP02878.
XX
XX 30-APR-1999; 99JP-0125169.
XX
XX (MATS/) MATSUMOTO A.
XX
XX Matsumoto A;
XX
XX WPI; 2000-687534/67.
XX
XX Human brain carboxypeptidase B isolated from the hippocampus useful for
XX screening agents for the treatment of Alzheimer's and other brain
XX disorders -
XX
XX Disclosure; Page 74-75; 84pp; Japanese.
XX
XX This invention describes a novel protein with peptidase activity
XX against brain beta-amyloid precursor protein which has been isolated from
XX human hippocampus and which has cerebroprotective, antialzheimers,
XX neurotropic, neuroprotective and hemostatic activity and which can be used
XX as a vaccine or for gene therapy. The protein, and compounds identified
XX by screening as promoters or inhibitors of its activity, are used to
XX regulate beta-amyloid accumulation in the brain and treat or prevent
XX diseases in which this occurs, such as Alzheimer's, senile dementia,
XX inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
XX Sequence 246 AA;
SQ
Query Match 74.4%; Score 1338; DB 21; Length 246;
Best Local Similarity 100.0%; Pred. No. 1,2e-131;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 93 ASASYEQVHSINETYSWIEFTEHRHPDLTKIHGSSPEKYLVLKSGKEQTAKNAI 152
DB 1 asasyeqvhsinetysewiftehphdltknhgsspekylvlksgkeqtaknai 60
OY 133 WIDCGIHAREWISPAFCIMFICGHNMRKNSFYANNHCIGTDLNRNFASKHWCCEGASS 212
DB 61 widcgiharewispaflcwficghnmwrknsfyannhncigdlrnrfaskhwccegass 120
OY 213 SCSSEYCGIYPESPEKAAVASFLRNINQIKAYISMSYSQHIVFPYSTRSKDH 272
DB 121 sscsetyqgiypesepkavassflrlninqikayismhsysqhlvfpysyrskskdhe 180
OY 273 ELISLVAEAVRAIEKTSKNTRTYHGHGSETLYLAPGGDDMYDGIKYSTSNPVEKL 332
DB 181 elslvaeavraiektskntrtyhghgsetlylapggddwlydglkystsnpvekl 240
OY 333 LPLSLK 338
DB 241 lplslk 246
RESULT 8
AAB58129
ID AAB58129 standard; Protein: 211 AA.
XX
XX AAB58129;
XX
XX 14-MAR-2001 (first entry)
XX
XX Lung cancer associated polypeptide sequence SEQ ID 467.
XX
XX Human: lung cancer associated protein; neuroprotective; cytostatic;
XX cardioactive; immunomodulatory; muscular active; vulnery;
XX gastrointestinal; nephrotropic; antineoplastic; gynecological;
XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX proliferative disorder; wound healing; infectious disease.
XX

```

```

OS Homo sapiens.
XX
XX WO200055180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000MO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
XX WPI; 2000-587514/75.
XX
XX N-PSDB; AAF18005.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
XX antigens, useful for treatment, prevention, and diagnosis of disorders
XX such as lung cancer -
XX
XX Claim 11; Page 944-945; 1425pp; English.
XX
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
XX associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX associated proteins and polynucleotide sequences; their agonists, and
XX antagonists may have neuroprotective; cytostatic; cardioactive;
XX immunomodulatory; muscular active general; vulnery; gastrointestinal
XX general; nephrotropic; antineoplastic; gynecological; or antibacterial
XX activity. The invention also includes antibodies specific for the
XX protein or polynucleotide sequences. The lung cancer associated
XX polynucleotide sequences may be used for detection of lung cancer,
XX chromosome identification, as chromosome markers, and for numerous other
XX diagnostic or research purposes. The proteins may be used to treat
XX disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders. The proteins may also be used in the treatment of wounds and
XX infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
XX peptide AAB58549 are used in the course of the invention for the
XX identification and characterization of the polynucleotide and protein
XX sequences.
XX
XX Sequence 211 AA;
SQ
Query Match 45.4%; Score 816.5; DB 21; Length 211;
Best Local Similarity 90.6%; Pred. No. 4.7e-77;
Matches 154; Conservative 1; Mismatches 6; Indels 9; Gaps 2;
OY 163 WISPAFCIMFL-----GHNRMWRKNSRFYANNHCIGTDLNRNFASKHWCCEGASS 214
DB 3 wis-paflcmflmmymwysvknrmwrknsfyannhncigdlrnrfaskhwccegass 61
OY 215 CSFTYCGIYPESPEKAAVASFLRNINQIKAYISMSYSQHIVFPYSTRSKDH 274
DB 62 csetyqgiypesepkavassflrlninqikayismhsysqhlvfpysyrskskdhe 121
OY 275 SLVAEAVRAIEKTSKNTRTYHGHGSETLYLAPGGDDMYDGIKYST 324
DB 122 slvaeavraiektskntrtyhghgsetlylapggddwlydglkyst 171
RESULT 9
AAR75131
ID AAR75131 standard; Protein: 404 AA.
XX
XX AAR75131;
XX
XX 13-MAR-1996 (first entry)
XX
XX Porcine Tyr-His-Met Procarboxypeptidase B.
XX

```



```

PR 16-AUG-1995; 95GB-0016810.
PR 23-DEC-1994; 94GB-0026192.
XX
XX (ZENE ) ZENECA LTD.
XX
PI Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW,
PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
PI Tarragona-Fiol A, Taylorson CJ;
XX
DR WPI; 1996-321650/32.
DR N-PSDB; AAT42506.
XX
PT Two component system for anti-tumour therapy - comprising targeting
PT moiety linked to mutated enzyme which can transform an
PT anti-neoplastic prodng
XX
XX Reference Example 18; Page 140-142; 182pp; English.
XX
CC A two-component system for anti-tumour therapy comprises a targeting
CC moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC prodng. The system is based on antibody directed enzyme prodng therapy
CC (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC be used. The present sequence is that of proHCPB with a pelB leader
CC sequence encoded by PIC11738.
XX
SQ Sequence 424 AA;

Query Match 35.5%; Score 637.5; DB 17; Length 424;
Best Local Similarity 36.0%; Pred. No. 8,6e-58;
Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps 6;

OY 1 FOSGVIALPRTSRQVYQNLITTYEIVLMQPVYADLIVKKQVHFPVNASDVNVA 60
DB 29 fgekvfrvnvedenhiniirelastqtdfkwpsvctqkphstvdfrvkaedltven 88
OY 61 HUNVSGIPSCVLLADVEDLIQOOISNDTVSPRASASYEOYHSLNDIYSWIEFTRHPD 120
DB 89 vlkqnelgykvvlismnlrvneagfdrv---ratgshyekykwetlswtqvatempa 145
OY 121 MTKIHIGSFEEKYPLYLVKVGSEKQTAKNAIWDGIIHAREWISPAFCIMFI----- 173
DB 146 llsrsvlgtfegralylkv-gkaeqnkpafimdcgfharwispafcqfvravrtly 204
OY 174 -----GH-----NBMWRKRSFANNHCITGDLNRNPAK 203
DB 205 greiqvtelldkldfylvplvlnldgylytwksrfrwktlrsthtgscclgtdpnrnf-da 263
OY 204 HMCERGAASSSCSEYVGLYPESEPEVKAVASFLRRNINQIKAYISMHSYQHIVPYSY 263
DB 264 gwcetagsrnpdecycpaaseketkaladfirnklsikyllshsgmmilypysy 323
OY 264 TMSKSDHEELVLASEAVRAIEKTSKNTRYTHGSEFLYLAPEGGDWITDGLIKYSF 323
DB 324 ayklgenaelnalakatvkel-aslhgtkytgyppattlpaagsddwaydqgltysf 382
OY 324 T 324
DB 383 L 383

RESULT 13
AAG6562
ID AAG6562 standard; Protein; 324 AA.
XX
XX AAG6562;
XX
XX 22-OCT-2001 (first entry)
XX
XX Bothops jararaca carboxypeptidase homologue protein.
XX

```

```

KW secreted metallocarboxypeptidase; immunomodulatory;
KW antiparkinsonian; nootropic; anticonvulsant; neuroprotective;
KW osteopathic; antiarthritic; cerebroprotective; vasotropic; vulnerary;
KW antibacterial; antiviral; antifungal; antiinflammatory; antidiabetic;
KW antineumatic; cytostatic; antiallergic; haemostatic; infection;
KW cancer.
XX
XX Bothops jararaca.
XX
XX WO200157265-A1.
XX
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-0503783.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 29-SEP-2000; 2000US-0676135.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Boyle BJ, Mize NK, Arterburn MC, Yeung G, Tang YT, Lau C;
XX Drmanac RT, Wang M, Chen L, Yang Y;
XX WPI; 2001-465705/50.
XX
XX Novel metallocarboxy peptidase-like polypeptide is useful in treating
XX neurodegenerative, immune, etc. disorders
XX
XX Example 3; Fig 1; 146pp; English.
XX
XX The invention relates to an isolated polypeptide comprising an amino
XX acid sequence of 374, 41, 14, 15, 41, 22, 15, 27, 14, 9, 18, 20, 15,
XX 354, 165, 374, 354, or 41 amino acids, fully defined in the
XX specification, or its translated protein coding portion, mature protein
XX coding portion, extracellular portion or active domain. The polypeptide
XX is a human secreted metallocarboxypeptidase-like polypeptide. It
XX is useful for treating mammalian diseases including:
XX neurodegenerative diseases such as Parkinson's or Alzheimer's diseases,
XX osteoporosis, osteoarthritis, disorders linked to reduced tissue
XX growth, Huntington's disease, amyotrophic lateral sclerosis, stroke,
XX reperfusion, wound healing, infectious diseases, immune and autoimmune
XX disorders such as severe combined immunodeficiency, rheumatoid
XX arthritis, guillain-Barre syndrome, insulin dependent diabetes
XX mellitus, graft-versus-host disease and allergy, cancer, and
XX haemophilia. The polypeptide is also useful in medical imaging, as an
XX immunogen to generate antibodies, as a nutritional source or
XX supplement, and as a contraceptive. The present sequence is
XX a carboxypeptidase homologue protein of Bothops jararaca.
XX
SQ Sequence 324 AA;

Query Match 35.4%; Score 636.5; DB 22; Length 324;
Best Local Similarity 42.0%; Pred. No. 7.1e-58;
Matches 126; Conservative 54; Mismatches 77; Indels 43; Gaps 5;

OY 62 LNVSGIPSCVLLADVEDLIQOOISNDTVSPRASASYEOYHSLNDIYSWIEFTRHPDM 121
DB 2 lqsgslnyellnldqvdirdqdnha---rtagynyekyswekdawadadanaenpsl 58
OY 122 LTRIHIGSFEEKYPLYLVKVGSEKQTAKNAIWDGIIHAREWISPAFCIMFI----- 173
DB 59 vsrlqigtftgfpmpllkv-gkpyvkkalfidcgharewispafcqfvravrtly 117
OY 174 -----GHNMWRKRSFANNHCITGDLNRNPAK 204
DB 118 ketimqljlnklldfylvplvlnldgyvywksrwmwktlsvnagstclgtdpnrnfdaa- 176
OY 205 WCEGASASSSCSEYVGLYPESEPEVKAVASFLRRNINQIKAYISMHSYQHIVPYSY 264
DB 177 wcsvgarnpcsetygskspeketkaladfirnrsltdqytlthsgmmilypysyl 236

```

OY 265 RSKSKDEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDWYDLGIRYSFT 324
 Db 237 ydltsnknklnslakealrel-kvligteyly9p9aaltlypaagsgddwaydgqilyafit 295

RESULT 14
 AAM41414
 ID AAM41414 standard; Protein: 716 AA.

XX AAM41414;

DT 02-JUN-1998 (first entry)

DE PreproHCPB-linker-humanised 806.077 VH.

KM Anti-CEA antibody; carcinoembryonic antigen; 806.077 Ab; cancer therapy;
 KM cancer diagnosis; complementarity determining region.

OS Synthetic.

XX WO9742329-A1.

PD 13-NOV-1997.

XX 29-APR-1997; 97WO-GB01165.

PR 14-FEB-1997; 97GB-0003103.
 PR 04-MAY-1996; 96GB-0009405.

PA (ZENEC) ZENECA LTD.

PI Copley CG, Edge MD, Emery SC;

DR WPI: 1997-558987/51.
 DR N-PSDB: AAV17331.

PT Anti-carcinoembryonic antigen antibody 806.077 Ab - used for
 PT diagnosis and therapy of cancer

XX Example 75; Page 164-187; 208pp; English.

XX This sequence is preproHCPB-linker-humanised 807.077 VH, which is
 CC an example of the antibody of the invention. The antibody is an anti-CEA
 CC (carcinoembryonic antigen) antibody (preferably 806.077 Ab). Host cells
 CC or transgenic organisms transformed with DNA encoding the antibody, are
 CC used to make the antibody or conjugate. The conjugate is used in a
 CC medicament suitable for intravenous administration. The conjugate can be
 CC used for cancer therapy, selectively killing tumour cells. The antibody
 CC can be used for in vivo or in vitro diagnosis of cancer.

SQ Sequence 716 AA;

Query Match 34.5%; Score 619.5; DB 18; Length 716;
 Best Local Similarity 35.2%; Pred. No. 1.6e-55;
 Matches 127; Conservative 69; Mismatches 122; Indels 43; Gaps 6;

OY 1 FOSGOVLALPRTRSOVYOLNLTTEYELVMOPVTADLYKKQVFFPNASVDNVKA 60

Db 22 fegekvtrvvedenhnllrelastqldfkwpsvltqkphstvfirkaedtlven 81

OY 61 HLANVSGIPCSVLADVDLIIQOQISNDTVSPRASASYEQVHSLNETYSMTIEFTTERHPD 120

Db 82 vlnqnelqykvllsnlnvveaqdldrv---ratghseyekynkwtleatvtqyatenpa 138

OY 121 MLRKIHGSSFEKRYPLVLYKVSKEQOTAKNAIWDICGIAHREWISPAFLMFI----- 173

Db 139 llsrsvlgttfefgralylklv-gkaggnkpaifmdcgfharewlspsfcqwtvreatvty 197

OY 174 -----GH-----NRMRKNSFPAANNHCTGIDLNRNPAK 203

Db 198 grelqvtellldkldfyvlpvlnldgylytwktsrftwktststhsaclygtqdpnrf-da 256

OY 204 HMCDEGASSSCSEYTCGLYPESEPEVKAVASFLRNINQIAVYISMHSQHIYFYSY 263
 Db 257 gnceigastnnpdecylcypaaseketalaadfrlnklsstlkayltllnsyqmmlypysy 316

OY 264 TRSKSKDEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDWYDLGIRYSF 323

Db 317 ayklgenaelnalakavkel-aslhgtkylypggactlypsagtsdwdydgqlyrsf 375

OY 324 T 324

Db 376 t 376

RESULT 15
 AAE00502
 ID AAE00502 standard; Protein: 437 AA.

XX AAE00502;

DT 19-JUN-2001 (first entry)

DE Human carboxypeptidase-related protein #5.

KM Human; carboxypeptidase-related protein; protease; gene therapy;
 KM drug screening; mental; biological disorder; medical disorder.

OS Homo sapiens.

XX WO200123588-A1.

PD 05-APR-2001.

XX 29-SEP-2000; 2000WO-US26876.

PR 29-SEP-1999; 99US-0156685.

PA (LEXI-) LEXICON GENETICS INC.

PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

DR WPI: 2001-266171/27.

DR N-PSDB: AAD03837.

PT New isolated human carboxypeptidase polynucleotide useful for
 PT generating antibodies, as reagents in diagnostic assays and for
 PT screening for compounds useful for treating mental, biological or
 PT medical diseases

PS Claim 2; Page 30-31; 36pp; English.

XX The present sequence is a novel human protein (NHP),
 CC carboxypeptidase-related protein. The carboxypeptidase-related
 CC proteins share structural similarity with animal proteases, and
 CC especially carboxypeptidase B or carboxypeptidase A. The
 CC carboxypeptidase-related protein cDNA is useful for the detection of
 CC mutant human carboxypeptidases, or inappropriately expressed human
 CC carboxypeptidases for the diagnosis of disease, and also as a
 CC therapeutic. It is also useful in drug screening, for generation of
 CC antibodies, as reagents in diagnostic assays, for the identification of
 CC other cellular gene products related to human carboxypeptidases, and as
 CC reagents in assays for screening for compounds that are useful for
 CC treating mental, biological or medical disorders and diseases. Nucleotide
 CC constructs encoding functional NHPs can also be used in gene therapy.

SQ Sequence 437 AA;

Query Match 34.0%; Score 612; DB 22; Length 437;

Best Local Similarity 35.4%; Pred. No. 4.3e-55;

Matches 129; Conservative 62; Mismatches 127; Indels 46; Gaps 7;

OY 1 FOSGOVLALPRTRSOVYOLNLTTEYELVMOPVTADLYV--KKQVFFPNASVDNV 58

Db 37 yagdkvlfirfpkteeeayalkkisyqlkvdlwgpssisyvsegvtvdvhlpgngstra--l 94

QY 59 KAHLNWVGIPCSVYLADVEDLIQOQISNDTVSPRASAS--YEQYHSLNEIYSNIEFITE 116

Db 95 laflgeanlqykvlliedlqktlekgsalhtqnrtrrsisgynevyhaleelqmwmlnk 154

QY 117 RHPDMLTKIHIGSFEKPYLYLVKVSKEQTAKNATWIDCGIHAREWISPAFCUMFT--- 173

Db 155 tbsgllhmfslgrsyegrsllflkl-grtrslkravldcglharewlgpaicqwtvkea 213

QY 174 -----GHNRMWRKNRSFYANNHCIGTDLNRN 199

Db 214 lltyksdpmrkmnlhlyfymvfnvdgyhfswtndrfwrktcrsrnsrfrcrgvdanrn 273

QY 200 FASKHWCCEGASSSCSEFYCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYSQHIVF 259

Db 274 wkvk-wcdegasmhpddkycgpfepesevkvannflrkhrkhraylsfhayaqmlly 332

QY 260 PYSTYRSKSKMDHELSIVASEAVRAIEKTSKNRRTYHGHSEFLYLAPGGDDMIYDLGI 319

Db 333 pyeykyatlipnfrcvesaaykavnaqlsv-ygvvryrygpastcllyvssgssmdwaykngl 391

QY 320 KYSP 323

Db 392 pyaf 395

Search completed: September 18, 2002, 16:45:13
Job time: 380 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:49:04 ; Search time 55.22 Seconds
(without alignments)
1058.897 Million cell updates/sec

Title: US-09-980-881-3

Perfect score: 1798
Sequence: 1 FQSGQVIALPRTSRQVQL.....IKYSFTSNPPEKLLPLSLK 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:**
1: sp_archaea:**
2: sp_bacteria:**
3: sp_fungi:**
4: sp_human:**
5: sp_invertebrate:**
6: sp_mammal:**
7: sp_mmc:**
8: sp_organelle:**
9: sp_phage:**
10: sp_plant:**
11: sp_rodent:**
12: sp_virus:**
13: sp_vertebrate:**
14: sp_unclassified:**
15: sp_virus:**
16: sp_bacteriophage:**
17: sp_archaea:**

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1784	99.2	360	4	O9P2Y6
2	1699.5	94.5	423	4	O15114
3	1694.5	94.2	423	4	O961Y4
4	1487.5	82.7	422	11	O9UJH6
5	1477.5	82.2	422	11	O9QZFO
6	1466.5	81.6	422	11	O9EOV9
7	916	50.9	198	4	O9WT18
8	697.5	38.8	416	13	O9PUF2
9	646.5	36.0	416	6	O9XSP3
10	637.5	35.5	417	4	O96B08
11	627	34.9	412	11	P97597
12	627	34.9	417	4	O96E94
13	537	29.9	279	11	O9CVD1
14	521.5	29.0	419	6	O9T855
15	514.5	28.6	417	4	O96ON3
16	514.5	28.6	417	4	O96A12

17	493.5	27.4	419	4	O9BS67	O9BS67 homo sapien
18	477.5	26.6	247	4	O96KZ9	O96KZ9 homo sapien
19	463	25.8	1192	5	O9W475	O9W475 drosophila
20	461.5	25.7	422	5	O9VLZ2	O9VLZ2 drosophila
21	443	24.6	584	5	O9TZH1	O9TZH1 caenorhabd
22	440	24.5	424	5	O9VL87	O9VL87 drosophila
23	426	23.7	440	5	O9W478	O9W478 drosophila
24	409	22.7	424	5	O9VS66	O9VS66 drosophila
25	408.5	22.6	419	5	O9VLZ1	O9VLZ1 drosophila
26	406.5	22.6	427	5	O9U9K2	O9U9K2 aedes aegypt
27	406.5	22.6	430	5	O9VL86	O9VL86 drosophila
28	403	22.4	323	5	O19121	O19121 caenorhabd
29	401	22.3	540	5	O9TZH2	O9TZH2 caenorhabd
30	398	22.1	453	5	O9V342	O9V342 drosophila
31	392	21.8	423	5	O61532	O61532 drosophila
32	391.5	21.8	581	5	O23318	O23318 caenorhabd
33	375	20.9	467	5	O9VCM8	O9VCM8 drosophila
34	372.5	20.7	455	5	O9TZC6	O9TZC6 caenorhabd
35	370	20.6	1430	5	O9VX86	O9VX86 drosophila
36	364	20.2	666	5	O9XU75	O9XU75 caenorhabd
37	361	20.1	415	5	O9VS64	O9VS64 drosophila
38	359	20.0	545	5	O9V3S6	O9V3S6 caenorhabd
39	357.5	19.9	400	5	O9VS67	O9VS67 drosophila
40	350.5	19.5	354	5	O9VS68	O9VS68 drosophila
41	350.5	19.5	418	5	O961J8	O961J8 drosophila
42	343	19.1	312	5	O9VXR3	O9VXR3 drosophila
43	339	18.9	315	4	O9NR19	O9NR19 homo sapien
44	331.5	18.4	161	13	O9OW12	O9OW12 brachydanio
45	328	18.2	418	3	O14418	O14418 metarhizium

ALIGNMENTS

RESULT 1
ID O9P2Y6 PRELIMINARY: PRT: 360 AA.
AC O9P2Y6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE CARBOXYPEPTIDASE B-LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL: AB011969; BAA90475.1; -;
DR HSSP: P00730; 2CTC.
DR MEROPS: M14.009; -;
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF002244; Propep_M14; 1.
DR Pfam: PF002246; Zn_carboxypept; 2.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR SEQUENCE 360 AA; 40935 MW; BF670B2F7A37C1CB CRC64;

Query Match 99.2%; Score 1784; DB 4; Length 360;
Best Local Similarity 99.1%; Pred. No. 4.8e-155;
Matches 335; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 FOSGOVLAALPRTSRQOVQVONLTTTTEIYVLMQPTADLIYKKQVHFVNASDVNDVNA 60
Db 23 FQSGOVLAAALPRTSRQOVQVONLTTTTEIYVLMQPTADLIYKKQVHFVNASDVNDVNA 82
QY 61 HUNVSGIPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNFIYSWIEFITERHPD 120
Db 83 HUNVSGIPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNFIYSWIEFITERHPD 142
QY 121 MLTKIHIGSFEEKYPLVYLKVSKEQOTAKNAIWDGIGHAREMISPAFLMFIIGH 180
Db 143 MLTKIHIGSFEEKYPLVYLKVSKEQOAKNAIWDGIGHAREMISPAFLMFIIGHNRMR 202
QY 181 KNRSPFANNHCIGTDLNRNFASKHNCCEGASSSCSEYTCGLYPSSEPVKAVASFLRRN 240
Db 203 KNRSTYANNHCIGTDLNRNFASKHNCCEGASSSCSEYTCGLYPSSEPEVKAVASFLRRN 262
QY 241 INQIKAYISMHSYQHIYEPYSTRSKSKDHRELVLVASEAVRALEKTSKNTRYHGHGS 300
Db 263 INQIKAYISMHSYQHIYEPYSTRSKSKDHRELVLVASEAVRALEKTSKNTRYHGHGS 322
QY 301 ETLYLAPGGGDMYIDLGIKISFTSNPVEKILPLSLK 338
Db 323 ETLYLAPGGGDMYIDLGIKISFTSNPVEKILPLSLK 360
RESULT 2
Q15114 PRELIMINARY; PRT; 423 AA.
ID Q15114
AC Q15114
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PCPB PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER.
RX MEDLINE=92042093; PubMed=1939207;
RA Epton D., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
J Biol. Chem. 266:21833-21838(1991).
DR EMBL: M75106; AAA60042.1; -.
DR HSP; P00730; IPT.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; PROPEP_M14.
DR InterPro; IPR000834; ZN_CARDOPEPT.
DR Pfam; PF02244; PROPEP_M14; 1.
DR Pfam; PF00246; ZN_CARDOPEPT; 1.
DR PRINTS; PR00765; CRBOXPTASEA.
FT CHAIN 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ SEQUENCE 423 AA; 48442 MW; 9B383272F6E79F4 CRC64;
Query Match 94.5%; Score 1699.5; DB 4; Length 423;
Best Local Similarity 89.8%; Pred. No. 3,3e-147;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 FOSGOVLAALPRTSRQOVQVONLTTTTEIYVLMQPTADLIYKKQVHFVNASDVNDVNA 60
Db 23 FQSGOVLAAALPRTSRQOVQVONLTTTTEIYVLMQPTADLIYKKQVHFVNASDVNDVNA 82
QY 61 HUNVSGIPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNFIYSWIEFITERHPD 120
Db 83 HUNVSGIPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNFIYSWIEFITERHPD 142
QY 121 MLTKIHIGSFEEKYPLVYLKVSKEQOTAKNAIWDGIGHAREMISPAFLMFIIGH 175
Db 143 MLTKIHIGSFEEKYPLVYLKVSKEQOTAKNAIWDGIGHAREMISPAFLMFIIGH 175

```

```

Db 143 MLTKIHIGSFEEKYPLVYLKVSKEQOTAKNAIWDGIGHAREMISPAFLMFIIGHTOFY 202
QY 176 -----NRMRKNSRSPYANNHCIGTDLNRNFASK 203
Db 203 GIIQYTNLRLVDFYVMPVNVVDGYDSMKRNRMRKNSRSPYANNHCIGTDLNRNFASK 262
QY 204 HMCCEGASSSCSEYTCGLYPSSEPEVKAVASFLRRNINQIKAYISMHSYQHIYEPY 263
Db 263 HMCCEGASSSCSEYTCGLYPSSEPEVKAVASFLRRNINQIKAYISMHSYQHIYEPY 322
QY 264 TRSKSDHRELVLVASEAVRAIEKTSKNTRYHGHSEETLYLAPGGGDMYIDLGIKYSF 323
Db 323 TRSKSDHRELVLVASEAVRAIEKTSKNTRYHGHSEETLYLAPGGGDMYIDLGIKYSF 382
QY 324 T 324
Db 383 T 383
RESULT 3
Q061Y4 PRELIMINARY; PRT; 423 AA.
ID Q061Y4
AC Q061Y4
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B2 (PLASMA).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKETAL MUSCLE;
RA Strausberg R.;
RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC007057; AAH07057.1; -.
SQ SEQUENCE 423 AA; 48412 MW; 9B383E03F6E7CF5 CRC64;
Query Match 94.2%; Score 1694.5; DB 4; Length 423;
Best Local Similarity 89.5%; Pred. No. 9,4e-147;
Matches 323; Conservative 0; Mismatches 1; Indels 37; Gaps 1;
QY 1 FOSGOVLAALPRTSRQOVQVONLTTTTEIYVLMQPTADLIYKKQVHFVNASDVNDVNA 60
Db 23 FQSGOVLAAALPRTSRQOVQVONLTTTTEIYVLMQPTADLIYKKQVHFVNASDVNDVNA 82
QY 61 HUNVSGIPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNFIYSWIEFITERHPD 120
Db 83 HUNVSGIPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNFIYSWIEFITERHPD 142
QY 121 MLTKIHIGSFEEKYPLVYLKVSKEQOTAKNAIWDGIGHAREMISPAFLMFIIGH 175
Db 143 MLTKIHIGSFEEKYPLVYLKVSKEQOAKNAIWDGIGHAREMISPAFLMFIIGHTOFY 202
QY 176 -----NRMRKNSRSPYANNHCIGTDLNRNFASK 203
Db 203 GIIQYTNLRLVDFYVMPVNVVDGYDSMKRNRMRKNSRSPYANNHCIGTDLNRNFASK 262
QY 204 HMCCEGASSSCSEYTCGLYPSSEPEVKAVASFLRRNINQIKAYISMHSYQHIYEPY 263
Db 263 HMCCEGASSSCSEYTCGLYPSSEPEVKAVASFLRRNINQIKAYISMHSYQHIYEPY 322
QY 264 TRSKSDHRELVLVASEAVRAIEKTSKNTRYHGHSEETLYLAPGGGDMYIDLGIKYSF 323
Db 323 TRSKSDHRELVLVASEAVRAIEKTSKNTRYHGHSEETLYLAPGGGDMYIDLGIKYSF 382
QY 324 T 324
Db 383 T 383

```

```
RESULT 4
ID 09JHH6 PRELIMINARY: PRT: 422 AA.
AC 09JHH6;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE R (THROMBIN-ACTIVATABLE FIBRINOLYSIS INHIBITOR)
DE (1110032P04RIK PROTEIN).
GN CPB2 OR TAFI OR 1110032P04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20341711; PubMed=10878383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA Campbell W., Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT whereas carboxypeptidase N is not.";
RL J. Immunol. 165:1053-1058(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20201966; PubMed=10739389;
RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
RT inhibitor.";
RL Thromb. Haemost. 83:297-303(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Sato T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nkaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Stadli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Williams L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AB021968; BAB03402.1; -
DR EMBL; AF164524; AAF62385.1; -
DR EMBL; AK004045; BAB3141.1; -
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -.
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Proper_M14.
DR InterPro; IPR000834; Zn_carboxept.
DR Pfam; PF02244; Proper_M14; 1.
DR Pfam; PF00246; Zn_carboxept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;
```

```
Query Match 82.7%; Score 1487.5; DB 11; Length 422;
Best Local Similarity 78.1%; Pred. No. 8,4e-128;
Matches 282; Conservative 20; Mismatches 22; Indels 37; Gaps 1;
```

1 FOSGOVLAALPRTSRQVOLLQNTTTEYEVLMQPTADLIYKKQVHFVNASDVNVA 60

```
Db 22 FOSGOVLAALPRTSRQVOLLQNTTTEYEVLMQPTAEFIEKKEVHFVNASDVSVNA 81
61 HLNVSGIPCSVLADVEDLIQOQISNDTVSPRASASYEQYHSLNTEYSWIEFTERRHPD 120
82 HLNVSRIPFVNLNNVEDLIEQOTFNDTVSPRASASYEQYHSLNTEYSWIEFTERRHPD 141
121 MLTKIHGSSFEKPYLYLVKSGKEQTAKNALWIDCGIHAREMISPAFLMGVYTOGH 175
142 MLQKITYGSSFEKPYLYLVKSGKEGRINKNALWIDCGIHAREMISPAFLMGVYTOGH 201
176 -----NRMRKNSFYANNHCIGTDLNRNPAK 203
202 GKENLYTRLLRHVDYFIMEVNVGYDTWKKRMRRKRNSAKNRNCVGTDLNRNPAK 261
204 HMCSEGASSSCSEYTCGLYSESEPEVKAVASFLRRNINOIKAYISMHSOHTVPSY 263
262 HMCSEGASSSCSEYTCGLYSESEPEVKAVADFLRRNIDIKAYISMHSOHTVPSY 321
264 TRSKSDHEELSLVASEAVRALEKTSKNTRYTHGSEETLYLAPGGDWIYDLGIRYSP 323
322 NRSKSDHEELSLVASEAVRALESINKNTRYTHGSGESLYLAPGGSDWIYDLGIRYSP 381
324 T 324
382 T 382
RESULT 5
ID 09QZF0 PRELIMINARY: PRT: 422 AA.
AC 09QZF0;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE U.
GN CPB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAF1J;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Sato T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nkaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Stadli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Williams L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AB021968; BAB03402.1; -
DR EMBL; AF164524; AAF62385.1; -
DR EMBL; AK004045; BAB3141.1; -
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -.
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Proper_M14.
DR InterPro; IPR000834; Zn_carboxept.
DR Pfam; PF02244; Proper_M14; 1.
DR Pfam; PF00246; Zn_carboxept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
SQ SEQUENCE 422 AA; 48832 MW; B53FFB09943954E CRC64;
```

```
Query Match 82.2%; Score 1477.5; DB 11; Length 422;
Best Local Similarity 77.6%; Pred. No. 6,9e-127;
Matches 280; Conservative 21; Mismatches 23; Indels 37; Gaps 1;
```

1 FOSGOVLAALPRTSRQVOLLQNTTTEYEVLMQPTADLIYKKQVHFVNASDVNVA 60

22 FOSGOVLAALPRTSRQVOLLQNTTTEYEVLMQPTAEFIEKKEVHFVNASDVSVNA 81

61 HLNVSGIPCSVLADVEDLIQOQISNDTVSPRASASYEQYHSLNTEYSWIEFTERRHPD 120

82 HLNVSRIPFVNLNNVEDLIEQOTFNDTVSPRASASYEQYHSLNTEYSWIEFTERRHPD 141

121 MLTKIHGSSFEKPYLYLVKSGKEQTAKNALWIDCGIHAREMISPAFLMGVYTOGH 175

```

Db 142 MLCITIGSSFEKPYLVKVGSKEDRIKNAIMIDGIIHAREWISPAFLMFGYTOFH 201
OY 176 -----NRMRKNSFYANNHCIGTDLNRFPASK 203
Db 202 GKENTYRLRLRHVDYIMPMNVNDGYDTWKKRBMKRRKRSNAKKNRCGTDLNRFPASK 261
OY 204 HMCCEGASSSCSECTCYGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIVFPYSY 263
Db 262 HMCCEGASSSCSECTCYGLYPESEPEVKAVADFLRRNIHIIKAYISMHSYQOILFPYSY 321
OY 264 TRSKSDHELSIVASEAVRAIEKTSKNTRYTHGSESTLYLAPGGDDWITDGIKYSF 323
Db 322 NRSKSDHELSIVASEAVRAIEKTSKNTRYTHGSESTLYLAPGGDDWITDGIKYSF 381
OY 324 T 324
Db 382 T 382

RESULT 6
O9EUV9 PRELIMINARY: PRT: 422 AA.
AC O9EUV9:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PRE-PROCARBOXYPEPTIDASE R.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20471387; PubMed=11021404;
RA Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA Okada N., Okada H.;
RT "Molecular cloning and partial characterization of rat
RT procarboxypeptidase R and carboxypeptidase N.";
RL Microbiol. Immunol. 44:719-728(2000).
DR HSP; AB042598; BAB18617.1; -.
DR HSP; P00730; ZCTC.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
KW Carboxypeptidase.
SQ SEQUENCE 422 AA; 48826 MW; FFFD32A51A9366C8 CRC64;

Query Match 81.6%; Score 1466.5; DB 11; Length 422;
Best Local Similarity 76.7%; Pred. No. 7e-126;
Matches 277; Conservative 22; Mismatches 25; Indels 37; Gaps 1;

OY 1 FOSGOVLAALPRTSRQOVQLONTLTTEYIYLQPVTAADLVKKQVHFFYNASDVNVKA 60
Db 22 FOSGHVLAALPRTSRQOVQLONTLTTEYIYLQPVTAADLVKKQVHFFYNASDVNVKA 81
OY 61 HLVNSGIPCSVLLADVEDLIQOQISNDTVSPRASASYEQYHSLNEIYSWIEFTERHPD 120
Db 82 YLVNSRIPEFVNLNNVDDLIQOQISNDTVSPRASASYEQYHSLNEIYSWIEFTERHPD 141
OY 121 MLTKIHGSSFEKPYLVKVGSKEDQAKNAIWDGIIHAREWISPAFLMFGI 175
Db 142 MLCITIGSSFEKPYLVKVGSKEDQAKNAIWDGIIHAREWISPAFLMFGI 201
OY 176 -----NRMRKNSFYANNHCIGTDLNRFPASK 203
Db 202 GKENTYRLRLRHVDYIMPMNVNDGYDTWKKRBMKRRKRSNAKKNRCGTDLNRFPASK 261
OY 204 HMCCEGASSSCSECTCYGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIVFPYSY 263

```

```

Db 262 HMCCEGASSSCSECTCYGLYPESEPEVKAVADFLRRNIHIIKAYISMHSYQOILFPYSY 321
OY 264 TRSKSDHELSIVASEAVRAIEKTSKNTRYTHGSESTLYLAPGGDDWITDGIKYSF 323
Db 322 NRSKSDHELSIVASEAVRAIEKTSKNTRYTHGSESTLYLAPGGDDWITDGIKYSF 381
OY 324 T 324
Db 382 T 382

RESULT 7
O9NT18 PRELIMINARY: PRT: 198 AA.
AC O9NT18:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BA139H14.2 (CARBOXYPEPTIDASE B2 (PLASMA)) (FRAGMENT).
GN CPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tromans A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL137141; CAB92622.1; -.
DR HSP; P00730; ICPX.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
KW Carboxypeptidase.
FT NON_TER 198
SQ SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;

Query Match 50.9%; Score 916; DB 4; Length 198;
Best Local Similarity 99.4%; Pred. No. 6.2e-76;
Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FOSGOVLAALPRTSRQOVQLONTLTTEYIYLQPVTAADLVKKQVHFFYNASDVNVKA 60
Db 23 FOSGOVLAALPRTSRQOVQLONTLTTEYIYLQPVTAADLVKKQVHFFYNASDVNVKA 82
OY 61 HLVNSGIPCSVLLADVEDLIQOQISNDTVSPRASASYEQYHSLNEIYSWIEFTERHPD 120
Db 83 HLVNSGIPCSVLLADVEDLIQOQISNDTVSPRASASYEQYHSLNEIYSWIEFTERHPD 142
OY 121 MLTKIHGSSFEKPYLVKVGSKEDQAKNAIWDGIIHAREWISPAFLMFGI 175
Db 143 MLCITIGSSFEKPYLVKVGSKEDQAKNAIWDGIIHAREWISPAFLMFGI 197

RESULT 8
O9PUF2 PRELIMINARY: PRT: 416 AA.
AC O9PUF2:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE HOMOLOG.
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubridae;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_Taxid=8724;
RN [1]
RP SEQUENCE FROM N.A.
RA Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;

```



```

Db 22 FBEKAFRRVNNEDENINIRELASTTQIDFMKPSVTOIKPHSTVDKRVKADVTVEN 81
QY 61 HLNVSIGIPCSVLLADVEDLIDQOISNDTVSPRASASYEQYHSNLEIYSWIEFITERHPD 120
Db 82 YKONELQKVLISLNLRNVEAOFDSRV---RATGSHYKYNKWEITEAMTQOVATENRA 138
QY 121 MLTKIHGSSFEKYPYLVLKYSKGETAKNAIWDGIGHAREWISPAFCWMT----- 173
Db 139 LISRSVIGTFEGRAIYLLKV-GKAGQNKPAJFMDCGFHAREWISPAFCQWVREAVRTY 197
QY 174 -----GH-----NRMKRRNSFYANNHCIGTDLNRNFASK 203
Db 198 GREIQVTELLDKIDFVLPVNLIDGIYTWTSKRFKRTKSTHTGSSCIGTDPNRNF-DA 256
QY 204 HMCBEGASSSCSEYTCGLYPESEPEVKAVASFLRRNIQKAYISMHSYQHIYFYSY 263
Db 257 GCEIGASHPNCDDEYCPAASESEKTKALADPFRNKLISSIKAYILHSYQSMILPYFSY 316
QY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMTYDLGIRYSF 323
Db 317 AKLCENNAELNALAKATVKEL-ASLHGTYKYTGPGATTYPPAAGSDOMAYDQIRYSF 375
QY 324 T 324
Db 376 T 376

```

```

RESULT 11
ID P97597 PRELIMINARY; PRT: 412 AA.
AC P97597;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MAST CELL CARBOXYPEPTIDASE A PRECURSOR (FRAGMENT).
GN R-CPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE DAWLEY;
RX MEDLINE=97149430; PubMed=8996238;
RA Lutzeltschab C., Pejler G., Aveskogh M., Hellman L.;
RT "Secretory granule proteases in rat mast cells. Cloning of 10
RT different serine proteases and a carboxypeptidase A from various rat
RT mast cell populations."
RL EMBL: U67914; AAB48267.1; -.
DR HSSP; P09955; INSA.
DR MEROPS; M14.010; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KM signal; Carboxypeptidase.
FT NON_TER 1
FT STGMN 1
FT CHAIN 11 412 MAST CELL CARBOXYPEPTIDASE A.
SQ SEQUENCE 412 AA; 47944 MW; DAB59555FC49137D CRC64;

```

```

Query Match 34.9%; Score 627; DB 11; Length 412;
Best Local Similarity 37.5%; Pred. No. 5e-49;
Matches 135; Conservative 55; Mismatches 128; Indels 42; Gaps 6;
QY 1 FOSGOVLAALPRTSRQOVQVQNTTTEYIYLMQPVTAADLVKKQYHFFVNASDVNVKA 60
Db 16 FDEKVFRRVVLQDEKQASLTKNLQTLTEIDFWPDAIHDAIVNMVDFRTERESQIOTS 75

```

```

QY 61 HLNVSIGIPCSVLLADVEDLIDQOISNDTVSPRASASYEQYHSNLEIYSWIEFITERHPD 120
Db 76 TLEQHKMDYEELLINLOEIDKQF--DYKEELAGHSHSAKYANDMKIYSWTEKWEKHE 133
QY 121 MLTKIHGSSFEKYPYLVLKYSKGETAKNAIWDGIGHAREWISPAFCWMT----- 173
Db 134 MYSRIKISTYEDNPNLYLKI-GRDGERKALFMDCGIHAHEWISPAFCQWVYQAAKSY 192
QY 174 GHN-----NRMKRRNSFYANNHCIGTDLNRNFASK 203
Db 193 GKNKIMTKLDRMNFYLPVFNVDGYIWSWTKRMRMRNRKSNKNSNCIGTDLNRNF-DV 251
QY 204 HMCBEGASSSCSEYTCGLYPESEPEVKAVASFLRRNIQKAYISMHSYQHIYFYSY 263
Db 252 SWDSSPNNDNCLSYKRPAPASESEKTKAVNFTSHLSNSTIKAYITTHYSQSMILPYFSY 311
QY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMTYDLGIRYSF 323
Db 312 TIKLPNHQDLKVAIATDVL-SSRYETRIYIGPIASTIKYTGSSSLDAWYDLGIRHTF 370

```

```

RESULT 12
ID Q96E94 PRELIMINARY; PRT: 417 AA.
AC Q96E94;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE SIMILAR TO CARBOXYPEPTIDASE A3 (MAST CELL).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW, AND ACUTE MYELOCYTIC LEUKEMIA;
RA Strimberg R.;
RA Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: BC012613; AAH12613.1; -.
DR Carboxypeptidase.
KM Carboxypeptidase.
SQ SEQUENCE 417 AA; 48714 MW; F781EEF23F570E98 CRC64;

```

```

Query Match 34.9%; Score 627; DB 4; Length 417;
Best Local Similarity 35.7%; Pred. No. 5.1e-49;
Matches 129; Conservative 71; Mismatches 117; Indels 44; Gaps 7;

```

```

QY 1 FOSGOVLAALPRTSRQOVQVQNTTTEYIYLMQPVTAADLVKKQYHFFVNASDVNVKA 60
Db 21 FDEKVFRRVRRPDEKQADIIKDLAKTNEIDFWPCTHHVAAKMAYDFRVSERESQAIOS 80
QY 61 HLNVSIGIPCSVLLADVEDLIDQOIS-NDTVSPRASASYEQYHSNLEIYSWIEFITERHPD 119
Db 81 ALDQNKMHYIILLHDQEELEKQFDVKEDIPGRHS---YAKINNMKEIYAWTEKMDKYP 137
QY 120 MLTKIHGSSFEKYPYLVLKYSKGETAKNAIWDGIGHAREWISPAFCWMT----- 173
Db 138 EMVSRIKIGSTYEDNPNLYLKI-GEKNERRKALFMDCGIHAHEWISPAFCQWVYQAAKTY 196
QY 174 -----GH-----NRMKRRNSFYANNHCIGTDLNRNFASK 202
Db 197 YGNKIMTKLDRMNFYLPVFNVDGYIWSWTKRMRMRNRKSNKNSNCIGTDLNRNF-N 255
QY 203 HMCBEGASSSCSEYTCGLYPESEPEVKAVASFLRRNIQKAYISMHSYQHIYFYSY 262
Db 256 ASWNSIPNTNDPCADNYSRAPSASESEKTKAVNFTSHLSNEIKVITTHYSQSMILPYFSY 315
QY 263 YTRSKDKHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMTYDLGIRYS 322
Db 316 YTSKLPNHNEDIAKAKIGTDVL-STRYETRIYIGPIESTIYPISSSLDAWYDLGIRHT 374
QY 323 F 323

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:45:13 ; Search time 65 Seconds

(without alignments)
420.372 Million cell updates/sec

Title: US-09-980-881-4

Perfect score: 1338

Sequence: 1 AASAYEQHSLNEITSWIE.....IKYFSTSPVEKILPLSLK 246

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Genesec_032802.*
1: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/genesec/genesecp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/genesec/genesecp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1338	100.0	246	21	Human brain carbox
2	1338	100.0	338	21	Human brain carbox
3	1325	99.0	360	21	Human brain carbox
4	1239.5	92.6	423	17	Human plasma carbox
5	1239.5	92.6	423	18	Human plasma carbox
6	1239.5	92.6	423	20	Human plasma carbox
7	1234.5	92.3	423	14	Human plasma carbox
8	816.5	61.0	211	21	Human plasma carbox
9	612.5	45.8	324	22	Human plasma carbox
10	610	45.6	307	17	Human plasma carbox
11	582	43.5	306	16	Human plasma carbox

12	582	43.5	306	22	AAU04477
13	582	43.5	329	17	AAU75131
14	567	42.4	349	17	AAU06174
15	567	42.4	349	17	AAU06173
16	567	42.4	349	17	AAU06172
17	567	42.4	349	17	AAU06171
18	567	42.4	349	17	AAU06170
19	560	41.9	349	17	AAU06169
20	560	41.9	349	17	AAU06168
21	559	41.8	349	17	AAU06167
22	559	41.8	349	17	AAU06166
23	556	41.6	349	18	AAU13750
24	554	41.4	349	18	AAU13752
25	554	41.4	349	18	AAU13751
26	553	41.3	349	18	AAU13750
27	552	41.3	349	18	AAU13753
28	551	41.2	349	18	AAU13754
29	551	41.2	349	18	AAU13755
30	551	41.2	349	18	AAU13756
31	551	41.2	349	18	AAU13757
32	550.5	41.1	354	22	AAU13758
33	550.5	41.1	354	22	AAU13759
34	550.5	41.1	354	22	AAU13760
35	550.5	41.1	354	22	AAU13761
36	550.5	41.1	354	22	AAU13762
37	549	41.0	349	18	AAU13763
38	549	41.0	349	18	AAU13764
39	549	41.0	349	18	AAU13765
40	549	41.0	349	18	AAU13766
41	549	41.0	349	18	AAU13767
42	549	41.0	349	18	AAU13768
43	548	41.0	349	18	AAU13769
44	546	40.8	349	18	AAU13770
45	492	36.8	399	22	AAU01664

ALIGNMENTS

RESULT 1	AAU1459	standard; Protein; 246 AA.
ID	AAU1459	standard; Protein; 246 AA.
XX	AAU1459	
AC	AAU1459	
DT	01-MAR-2001 (first entry)	
XX	01-MAR-2001 (first entry)	
DE	Human brain carboxypeptidase B protein SEQ ID NO 4.	
XX	Human brain carboxypeptidase B protein SEQ ID NO 4.	
KW	Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase; treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine; cerebroprotective; antialzheimer's; neurotrophic; neuroprotective;	
KW	hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage; Down's syndrome; head trauma.	
XX	Down's syndrome; head trauma.	
OS	Homo sapiens.	
XX	Homo sapiens.	
PN	WO200066717-A1.	
XX	WO200066717-A1.	
PD	09-NOV-2000.	
XX	09-NOV-2000.	
PF	01-MAY-2000; 2000WO-JP02878.	
XX	01-MAY-2000; 2000WO-JP02878.	
PR	30-APR-1999; 99JP-0125169.	
XX	30-APR-1999; 99JP-0125169.	
PA	(MATS/) MATSUMOTO A.	
XX	(MATS/) MATSUMOTO A.	
PI	Matsumoto A.	
XX	Matsumoto A.	
DR	WPI: 2000-687534/67.	
XX	WPI: 2000-687534/67.	
PT	Human brain carboxypeptidase B isolated from the hippocampus useful for screening agents for the treatment of Alzheimer's and other brain	

PT disorders -
 PS Disclosure; Page 74-75; 84pp; Japanese.
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity, and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 CC
 SQ Sequence 246 AA;

Query Match 100.0%; Score 1338; DB 21; Length 246;
 Best Local Similarity 100.0%; Pred. No. 5.3e-134;
 Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASASYEYHSLNELYSWIEFTTERHPDMLTKIHGSSFEKYPYLKVGSKETAKNAI 60
 DB 1 ASASYEGYHSLNELYSWIEFTTERHPDMLTKIHGSSFEKYPYLKVGSKETAKNAI 60
 QY 61 WIDGCIHAREWISPAFCIMFIGNHMRKNSFYANNHCIGTDLNRNFASKHWCCEGASS 120
 DB 61 WIDGCIHAREWISPAFCIMFIGNHMRKNSFYANNHCIGTDLNRNFASKHWCCEGASS 120
 QY 121 SSCSEYTCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSOHIVFPYSTRSKSDHE 180
 DB 121 SSCSEYTCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSOHIVFPYSTRSKSDHE 180
 QY 181 ELISIVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMIYDLGKYSFTSNPVEK 240
 DB 181 ELISIVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMIYDLGKYSFTSNPVEK 240
 QY 241 LPLSLK 246
 DB 241 LPLSLK 246

RESULT 2
 ID AAB11458 standard; Protein; 338 AA.
 AC AAB11458;

DT 01-MAR-2001 (first entry)

DE Human brain carboxypeptidase B protein SEQ ID NO 3.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KM treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KM cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
 KM hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KM Down's syndrome; head trauma.

XX Homo sapiens.

OS

PN WO200006717-A1.

PD 09-NOV-2000.

PF 01-MAY-2000; 2000MO-JP02878.

PR 30-APR-1999; 99JP-0125169.

PA (MATS/) MATSUMOTO A.

PI Matsumoto A;

XX WPI; 2000-687534/67.

XX Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 PS Disclosure; Page 71-73; 84pp; Japanese.

CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity, and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 CC
 SQ Sequence 338 AA;

Query Match 100.0%; Score 1338; DB 21; Length 338;
 Best Local Similarity 100.0%; Pred. No. 8.6e-134;
 Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASASYEYHSLNELYSWIEFTTERHPDMLTKIHGSSFEKYPYLKVGSKETAKNAI 60
 DB 93 ASASYEGYHSLNELYSWIEFTTERHPDMLTKIHGSSFEKYPYLKVGSKETAKNAI 152
 QY 61 WIDGCIHAREWISPAFCIMFIGNHMRKNSFYANNHCIGTDLNRNFASKHWCCEGASS 120
 DB 153 WIDGCIHAREWISPAFCIMFIGNHMRKNSFYANNHCIGTDLNRNFASKHWCCEGASS 212
 QY 121 SSCSEYTCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSOHIVFPYSTRSKSDHE 180
 DB 213 SSCSEYTCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSOHIVFPYSTRSKSDHE 272
 QY 181 ELISIVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMIYDLGKYSFTSNPVEK 240
 DB 273 ELISIVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMIYDLGKYSFTSNPVEK 332
 QY 241 LPLSLK 246
 DB 333 LPLSLK 338

RESULT 3
 ID AAB11457 standard; Protein; 360 AA.
 AC AAB11457;

DT 01-MAR-2001 (first entry)

DE Human brain carboxypeptidase B protein.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KM treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KM cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
 KM hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KM Down's syndrome; head trauma.

XX Homo sapiens.

OS

PN WO200006717-A1.

PD 09-NOV-2000.

PF 01-MAY-2000; 2000MO-JP02878.

PR 30-APR-1999; 99JP-0125169.

PA (MATS/) MATSUMOTO A.

PI Matsumoto A;
XX WPI: 2000-687534/67.
DR N-PSDB: AAC81962.
XX
PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Claim 1; Page 68-71; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimer's,
CC neurotropic, neuroprotective and hemostatic activity, and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 360 AA;

Query Match 99.0%; Score 1325; DB 21; Length 360;
Best Local Similarity 98.8%; Pred. No. 2.3e-132;
Matches 243; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 ASASYEQYHSLNEIYSWTEFTEHHPDMLTKIHGSSPEKYPVLYLAKSGKEQTAKNAI 60
DB 115 ASASYEQYHSLNEIYSWTEFTEHHPDMLTKIHGSSPEKYPVLYLAKSGKEQTAKNAI 174
OY 61 WIDCGIHAREWISPAFCCLMFICGHNRMWRKNSFYANNHCIGTDLNRNFAASKHCEGAS 120
DB 175 WIDCGIHAREWISPAFCCLMFICGHNRMWRKNSFYANNHCIGTDLNRNFAASKHCEGAS 234
OY 121 SSCSTYGCGLYSEPEPKAVASFLRRNINQKATISMHSYQHIYFYSYTRSKSKDHE 180
DB 235 SSCSTYGCGLYSEPEPKAVASFLRRNINQKATISMHSYQHIYFYSYTRSKSKDHE 294
OY 181 ELTLVAEAVRAIEKTSKMTKRTYHGHSEFTLAPGCGDDWLYDGIKSPFNPPVEKL 240
DB 295 ELTLVAEAVRAIEKTSKMTKRTYHGHSEFTLAPGCGDDWLYDGIKSPFNPPVEKL 354
OY 241 LPLSLK 246
DB 355 LPLSLK 360

RESULT 4
AAR90293
ID AAR90293 standard; Protein; 423 AA.
XX
AC AAR90293;
XX
DT 12-APR-1996 (first entry)
XX
DE Human plasma carboxypeptidase B.
XX
KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Active-site 181
FT /note- "Residue suspected of being involved in
FT catalytic activity of enzyme."
FT Binding-site 183
FT /note- "Residue suspected of being part of the
FT binding site of enzyme."
FT Binding-site 256..257
FT /note- "Residues suspected of being part of the

FT binding site of enzyme."
FT Active-site 310
FT /note- "Residue suspected of being involved in
FT catalytic activity of enzyme."
FT Binding-site 312
FT /note- "Residue suspected of being part of the
FT binding site of enzyme."
FT Active-site 369
FT /note- "Residue suspected of being involved in
FT catalytic activity of enzyme."
FT Binding-site 370..371
FT /note- "Residues suspected of being part of the
FT binding site of enzyme."
FT Binding-site 394
FT /note- "Residue suspected of being part of the
FT binding site of enzyme."
PN US5474901-A.
XX
XX 12-DEC-1995.
PD
XX
XX 01-FEB-1991; 91US-0649591.
PF
XX
XX 01-FEB-1991; 91US-0649591.
PR
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
XX
XX
PA (GENTEC) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
PI WPI: 1996-039508/04.
DR N-PSDB: AAT11671.
XX
PT Antibody to human plasma carboxypeptidase B - useful for detecting
PT and purifying hPCPB for use in treating clotting disorders e.g.
PT haemophilia A
XX
XX
PS Disclosure; Figure 4; 40pp; English.
XX
CC An antibody which specifically binds human plasma carboxypeptidase B
CC (hPCPB) and does not cross react with other carboxypeptidases is
CC useful for the detection of hPCPB in vitro. The antibody is also
CC used for purifying hPCPB from a sample. Purification comprises
CC passing a sample thought to contain hPCPB over either a column to
CC which antibody has been bound, or a plasminogen affinity column,
CC eluting the column and then recovering the fraction containing the
CC hPCPB.
CC
SQ Sequence 423 AA;

Query Match 92.6%; Score 1239.5; DB 17; Length 423;
Best Local Similarity 86.2%; Pred. No. 3.7e-123;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

OY 1 ASASYEQYHSLNEIYSWTEFTEHHPDMLTKIHGSSPEKYPVLYLAKSGKEQTAKNAI 60
DB 115 ASASYEQYHSLNEIYSWTEFTEHHPDMLTKIHGSSPEKYPVLYLAKSGKEQTAKNAI 174
OY 61 WIDCGIHAREWISPAFCCLMFICGHNRMWRKNSFYANNHCIGTDLNRNFAASKHCEGAS 120
DB 175 WIDCGIHAREWISPAFCCLMFICGHNRMWRKNSFYANNHCIGTDLNRNFAASKHCEGAS 234
OY 181 ELTLVAEAVRAIEKTSKMTKRTYHGHSEFTLAPGCGDDWLYDGIKSPFNPPVEKL 240
DB 295 ELTLVAEAVRAIEKTSKMTKRTYHGHSEFTLAPGCGDDWLYDGIKSPFNPPVEKL 354
OY 241 LPLSLK 246
DB 355 LPLSLK 360

RESULT 4
AAR90293
ID AAR90293 standard; Protein; 423 AA.
XX
AC AAR90293;
XX
DT 12-APR-1996 (first entry)
XX
DE Human plasma carboxypeptidase B.
XX
KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Active-site 181
FT /note- "Residue suspected of being involved in
FT catalytic activity of enzyme."
FT Binding-site 183
FT /note- "Residue suspected of being part of the
FT binding site of enzyme."
FT Binding-site 256..257
FT /note- "Residues suspected of being part of the

DR N-PSDB: AAV74302.
 XX
 PT Detecting new polymorphism of human plasma carboxypeptidase B -
 PT comprises Alanine or Threonine at position 147 of protein by DNA or
 PT protein analysis, useful to detect risk of thrombotic disease in
 PT humans
 XX
 PS Disclosure: Page 25-26; 35pp; English.
 XX
 CC This sequence is the human plasma carboxypeptidase B (PCPB) mutant
 CC hPCBhtr147. The invention relates to a method for determining the
 CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
 CC comprises obtaining a prepared tissue or blood sample and determining the
 CC presence of DNA coding for naturally occurring polymorphs of the protein
 CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
 CC respectively). Determination of the relative distribution of the PCPB
 CC polymorphs in a patient's blood by genetic or protein analysis by the
 CC methods is useful to determine the risk of thrombotic disease in humans.
 CC Such assessments may be made by accumulating information concerning the
 CC relative distribution of the different polymorphs within the general
 CC population compared with populations known to be at risk and establishing
 CC a PCPB polymorph profile for at-risk patients.
 CC
 XX
 SQ Sequence 423 AA:
 Query Match 92.6%; Score 1239.5; DB 20; Length 423;
 Best Local Similarity 86.2%; Pred. No. 3.7e-123;
 Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
 OY 1 ASASYEYQHSLENIYSWIEFTEHPDMLTKIHGSSFEKYPYLVKYGKQOTAKNAI 60
 DB 115 asasyeqyhslnelyswieftehpdmlltkihgssfekyplylvkysgkeqtaknai 174
 OY 61 WTIDCGIHAREWISPAFLCWFIGH----- 83
 DB 175 wldcgiharewispafclwfighltcfygllygqyltlllrvdympvvnvdgdygswk 234
 OY 84 NRMWRKNSRFYANNHCIGDNLNRNFASKHMCCEGASSSCSEFYCGLYPSEPEPKAVAS 143
 DB 235 nrmwrknsrfyannhncigdlrnrfaskhmcceegassscsefycglypsepevkavas 294
 OY 144 FLRRNINQIKATISMHSYSGOHVFPYSTRSKSKDHEELVASEAVRAIEKTSKTRT 203
 DB 295 flrrninqikayismhsyqghvfpysftrskskdheelsvaseavraiektksktrlyt 354
 OY 204 HGHGSETLYAPGCGDDWYIDGIRKYSFT 232
 DB 355 hghgsetlylapgpgddwlyldgirkysft 383
 RESULT 7
 AAR36273
 ID AAR36273 standard; Protein: 423 AA.
 AC AAR36273;
 XX 24-AUG-1993 (first entry)
 XX Human plasma carboxypeptidase B.
 DE PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
 KM
 XX Homo sapiens.
 OS
 XX
 FH Key
 FT Peptide 1..22 Location/Qualifiers
 FT Cleavage-site //label= signal_peptide
 FT //note= "putative site of cleavage by trypsin to
 FT Active-site 181 activate PCPB as a carboxypeptidase"
 FT //note= "expected to be involved in catalytic

FT Binding-site 183 activity"
 FT /note= "expected to be involved in substrate
 FT binding"
 FT Active-site 184
 FT /note= "expected to be involved in catalytic
 FT activity"
 FT Binding-site 256..257
 FT /note= "expected to be involved in substrate
 FT binding"
 FT Active-site 310
 FT /note= "expected to be involved in catalytic
 FT activity"
 FT Binding-site 312
 FT /note= "expected to be involved in substrate
 FT binding"
 FT Active-site 369
 FT /note= "expected to be involved in catalytic
 FT activity"
 FT Binding-site 370..371
 FT /note= "expected to be involved in substrate
 FT binding; Asp 370 is the residue which
 FT determines specificity of the PCPB as
 FT a carboxypeptidase B"
 FT Binding-site 394
 FT /note= "expected to be involved in substrate
 FT binding"
 FT US5206161-A.
 XX 27-APR-1993.
 PD 01-FEB-1991; 91US-0649591.
 PF 01-FEB-1991; 91US-0649591.
 PR 01-FEB-1991; 91US-0649591.
 PA (GETH) GENENTECH INC.
 XX Drayna DT, Eaton DL;
 PI WPI: 1993-151724/18.
 DR N-PSDB; AA041001.
 XX
 PT New human plasma carboxypeptidase B - used as haemostatic
 PT regulator for clotting blood, partic. for treating blood clotting
 PT disorders, e.g. haemophilia
 XX
 PS Claim 2; Fig 4; 40pp; English.
 XX
 CC This amino acid sequence was deduced from a clone isolated from a
 CC human liver cDNA library. There is about 40% sequence identity
 CC between the deduced preproPCPB amino acid sequence and pro-human
 CC mast cell carboxypeptidase A and between preproPCPB and prepro-rat
 CC carboxypeptidase B. Human PCPB has the same substrate binding sites
 CC as, and shares 6 cysteine residues which form 3 intramolecular
 CC disulphide bonds with, bovine and rat carboxypeptidase B. The
 CC presence of the same amino acid (Asp at position 348 of the mature
 CC protein) at the region in carboxypeptidases that determines
 CC substrate specificity, suggests that PCPB represents a plasma-
 CC derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
 CC tPA of plasminogen to plasmin in the presence of fibrinogen.
 CC
 XX
 SQ Sequence 423 AA:
 Query Match 92.3%; Score 1234.5; DB 14; Length 423;
 Best Local Similarity 85.9%; Pred. No. 1.3e-122;
 Matches 231; Conservative 0; Mismatches 1; Indels 37; Gaps 1;
 OY 1 ASASYEYQHSLENIYSWIEFTEHPDMLTKIHGSSFEKYPYLVKYGKQOTAKNAI 60
 DB 115 asasyeqyhslnelyswieftehpdmlltkihgssfekyplylvkysgkeqtaknai 174

QY 61 WTDCGIAHREWISPAFLMFTGH----- 83
 DB 175 Wldcgihrewwspafclwflghlclqfyglqylnllrvdlympvvnvdygywkw 234
 QY 84 NRMNRKNSRYANNHCIGTDLNRNFASKHMCCEGASSSCSPTYGGLPESPEPKAVAS 143
 DB 235 nrmwknrsfyannhcltdlnrnfaskhmcceagassscetlyglpesepekavas 294
 QY 144 FLRRNINIKAVISMHSYQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
 DB 295 flrrlnqikayismhsyqhlvfpystrskskdheelslvaseavraiektsknttryt 354
 QY 204 HGHSEETLYLAPGGDDWYDLGIRYSFT 232
 DB 355 hghsetlylapggddwlydlylkyfst 383

RESULT 8
 AAB58129
 ID AAB58129 standard; Protein: 211 AA.
 AC AAB58129;
 DE 14-MAR-2001 (first entry)

Lung cancer associated polypeptide sequence SEQ ID 467.

Human: Lung cancer associated protein; neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular active; vulnery; gastrointestinal; nephrotoxic; antiinfective; gynecological; antibacterial; diagnosis; neural disorder; immune disorder; reproductive; proliferative disorder; wound healing; infectious disease.

OS Homo sapiens.
 PN W020005180-A2.
 PD 21-SEP-2000.
 PF 08-MAR-2000; 2000MO-US05918.
 PR 12-MAR-1999; 9905-0124270.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 PI Ruben SM;
 DR WPI; 2000-587514/55.
 DR N-PSDB; AAF18005.
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 PS Claim 11; Page 944-945; 1425pp; English.
 XX

Polynucleotide sequences AAF1982 - AAF18424 encode human lung cancer associated proteins represented in AAB58106 - AAB58548. Lung cancer associated proteins and polynucleotide sequences, their agonists, and antagonists may have neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular active general; vulnery; gastrointestinal general; nephrotoxic; antiinfective; gynecological; or antibacterial activity. The invention also includes antibodies specific for the protein or polynucleotide sequences. The lung cancer associated polynucleotide sequences may be used for detection of lung cancer, chromosome identification, as chromosome markers, and for numerous other diagnostic or research purposes. The proteins may be used to treat disorders such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, and proliferative disorders. The proteins may also be used in the treatment of wounds and infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are used in the course of the invention for the

CC Identification and characterisation of the polynucleotide and protein
 CC sequences.
 CC XX
 SQ Sequence 211 AA;

Query Match 61.0%; Score 816.5; DB 21; Length 211;
 Best Local Similarity 90.6%; Pred. No. 1.5e-78;
 Matches 154; Conservative 1; Mismatches 6; Indels 9; Gaps 2;

QY 71 WISPAFLMFT-----GNNRMRKNSRYANNHCIGTDLNRNFASKHMCCEGASSS 122
 DB 3 wis-mlcwlmwmyunyswknrmrknrsfyannhcltdlnrnfaskhmcceagasss 61

QY 123 CSFTYCGLPESPEPEKAVASFLRRNINOTKAYISMHSYQHIVFPYSYTRSKSKDHEEL 182
 DB 62 csetlycglypesepekavasflrrlnqikayismhsyqhlvfpystrskskdheel 121

QY 183 SLVASEAVRAIEKTSKNTRYTHGHSEETLYLAPGGDDWYDLGIRYSFT 232
 DB 122 slvaseavraiektsknttrythghsetlylapggddwlydlylkyfst 171

RESULT 9
 AAG6562
 ID AAG6562 standard; Protein: 324 AA.
 AC AAG6562;
 DE 22-OCT-2001 (first entry)

Bothops jararaca carboxypeptidase homologue protein.

XX secreted metalloproteinase; immunomodulatory;
 KW antiparkinsonian; nootropic; anticonvulsant; neuroprotective;
 KW osteopathic; antiarthritic; cerebroprotective; vasotropic; vulnery;
 KW antibacterial; antiviral; antitumor; antiinflammatory; antidiabetic;
 KW antithematic; cytostatic; antiallergic; haemostatic; infection;
 KW cancer.
 XX
 OS Bothops jararaca.
 PN W0200157265-A1.
 PD 09-AUG-2001.
 PF 05-FEB-2001; 2001MO-US03783.
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 29-SEP-2000; 2000US-0676135.
 PA (HYSE-) HYSEQ INC.
 PI Boyle BJ, Mize NK, Arterburn MC, Yeung G, Tang YT, Liu C;
 PI Dmanac RT, Wang M, Chen L, Yang Y;
 DR WPI; 2001-465705/50.
 XX

Novel metalloproteinase-like polypeptide is useful in treating neurodegenerative, immune, etc. disorders -
 PS Example 3; Fig 1; 146pp; English.
 XX

The invention relates to an isolated polypeptide comprising an amino acid sequence of 374, 41, 15, 41, 22, 15, 27, 14, 9, 18, 20, 15, 354, 165, 374, 354, or 41 amino acids, fully defined in the specification, or its translated protein coding portion, mature protein coding portion, extracellular portion or active domain. The polypeptide is a human secreted metalloproteinase-like polypeptide. It is useful for treating mammalian diseases including: neurodegenerative diseases such as Parkinson's or Alzheimer's diseases, osteoporosis, osteoarthritis, disorders linked to reduced tissue

CC growth, Huntington's disease, amyotrophic lateral sclerosis, stroke,
 CC reperfusion, wound healing, infectious diseases, immune and autoimmune
 CC disorders such as severe combined immunodeficiency, rheumatoid
 CC arthritis, Guillain-Barre syndrome, insulin dependent diabetes
 CC mellitus, graft-versus-host disease and allergy, cancer, and
 CC haemophilia. The polypeptide is also useful in medical imaging, as an
 CC immunogen to generate antibodies, as a nutritional source or
 CC supplement, and as a contraceptive. The present sequence is
 CC a carboxypeptidase homologue protein of Bothrops jararaca.

Sequence 324 AA:

Query Match 45.8%; Score 612.5; DB 22; Length 324;
 Best Local Similarity 45.0%; Pred. No. 1.5e-56;
 Matches 121; Conservative 44; Mismatches 63; Indels 41; Gaps 5;

OY 2 SASY-YEYQHSNLEIYSWIFETERHPDLTKIHGSSFEKPYLYLYKYSGEQTKAKNI 60
 Db 30 tagyryekynswekldawladanepsrlqtltegrpmplkv-gkpyvnhkai 88
 OY 61 WIDCGIHAREWISPAFCLMFI-----GH 83
 Db 89 fdcgfharewispacqwfveavrtlygketimqlnkldfylvlnldgyvyswkg 148
 OY 84 NRMKRNRFYANNHCIGTDLNRFASKHWCCEGASSSCSEYTCGLYPESEPEYKAVAS 143
 Db 149 smwkrtrstmasgscldgvdprnrf-wcsvgsarmpcsetycgskpesekekatalad 207
 OY 144 FLRRINQIKATISMHSYQHIYFPYSTRSKKDHEELSLVASEVRAIEKTSKTRT 203
 Db 208 flrrnrlstikaytlthsqmlylpydykldpenyealnalkgaekel-atlhgtkyt 266
 OY 204 HGHSELYIAPGSGDDWIYDLGIKYSFT 232
 Db 267 ypggaatltpaaggsddawdydgikysft 295

RESULT 10

AAW00602
 ID AAW00602 standard; Protein; 307 AA.

AC AAW00602;
 DT 09-NOV-1996 (first entry)
 DE Rat carboxypeptidase B mature enzyme.
 KM Carboxypeptidase-B; pro-enzyme; protease; insulin.
 OS Rattus sp.
 PN W09623064-A1.
 PD 01-AUG-1996.
 PP 25-JAN-1996; 96MO-US00995.
 PR 25-JAN-1995; 95US-0378233.
 PA (BIOT-) BIOTECHNOLOGY GEN CORP.
 PI Fulga N, Gorecki M, Hartman J, Mendelovitch S;
 DR WPI; 1996-362688/36.
 DR N-PSDB; AAF35760.
 PT Purified active recombinant carboxypeptidase B prodn. - by
 PT cleavage to give active form, used e.g. for insulin prodn.
 PS Example 1; Page 37-38; 49pp; English.

CC The amino acid sequence of the rat mature carboxypeptidase-B
 CC (CPB) (AAW00602) was deduced from a cDNA sequence (see also AAT35760)
 CC obtd. by PCR amplification of Sprague-Dawley rat pancreas cDNA.
 CC The sequence of the activation peptide (AAW00601) of pro-CPB has
 CC also been deduced. Pro-CPB (pref. with a C290S mutation) can be
 CC expressed in host cells, e.g. Escherichia coli, and subsequently
 CC recovered, refolded and cleaved with trypsin to yield the active
 CC enzyme. CPB produced this way is cheaper than porcine pancreatic
 CC enzyme, and is free of other proteases.

Sequence 307 AA:

Query Match 45.6%; Score 610; DB 17; Length 307;
 Best Local Similarity 45.0%; Pred. No. 2.6e-56;
 Matches 121; Conservative 43; Mismatches 65; Indels 40; Gaps 5;

OY 1 ASASVYEQHSNLEIYSWIFETERHPDLTKIHGSSFEKPYLYLYKYSGEQTKAKNI 60
 Db 1 asghsytkynwelleawliqvaldnpdlvtgsvlgtfegrmyviki-gktrpkpai 59
 OY 61 WIDCGIHAREWISPAFCLMFI-----GH 83
 Db 60 fdcgfharewispacqwfveavrtlyngelhmklldeldfylvlnldgyvvtwk 119
 OY 84 NRMKRNRFYANNHCIGTDLNRFASKHWCCEGASSSCSEYTCGLYPESEPEYKAVAS 143
 Db 120 dmwkrtrstmasgscldgvdprnrf-nagwcvegsarmpcsetycgskpesekekatalad 178
 OY 144 FLRRINQIKATISMHSYQHIYFPYSTRSKKDHEELSLVASEVRAIEKTSKTRT 203
 Db 179 flrrnrlstikaytlthsqmlylpydykldpenyealnalkgaekel-atlhgtkyt 237
 OY 204 HGHSELYIAPGSGDDWIYDLGIKYSFT 232
 Db 238 ypggaatltpaaggsddawdydgikysft 266

RESULT 11

AAR75132
 ID AAR75132 standard; Protein; 306 AA.

AC AAR75132;
 DT 13-MAR-1996 (first entry)
 DE Porcine carboxypeptidase B.
 KM Procarboxypeptidase B; carboxypeptidase B; Pichia; pCPB;
 KM human serum albumin; premating factor alpha; mating factor alpha;
 OS Sus scrofa.
 PN W09514096-A1.
 PD 26-MAY-1995.
 PP 16-NOV-1994; 94MO-US13142.
 PR 16-NOV-1993; 93US-0153258.
 PA (ELIL) LILLY & CO ELI.
 PI Fagerman JT, Greenen DP, Hersberger CL, Larson JL;
 PI Sterner JL, Zhang H;
 DR WPI; 1995-200386/26.
 DR N-PSDB; AAQ90601.
 PT DNA encoding porcine carboxypeptidase B - used for transforming
 PT host cells, partic. Pichia species, for prodn. of the enzyme

PS Example 1; Page 18; 34pp; English.
XX

The porcine carboxypeptidase B coding sequence can be placed in a bacterial or pre-f. Pichia yeast expression vector. The expression vector further comprises the signal peptide of either human serum albumin (designated pGCD3 - NRRL B-21029), premating factor alpha (designated pF4489 - NRRL B-21028), mating factor alpha (designated pF4474 - NRRL B-21032), or the porcine proCB signal peptide (designated pGCD27 - NRRL B-21027). The method can be used for producing large amounts of porcine carboxypeptidase B and when produced in Pichia yeast, the protein does not need solubilisation or folding. The produced enzyme is then used for pre-f. cleaving basic residues from the carboxyterminus of proteins.

sq	sequence	306 AA;
----	----------	---------

Query Match	43.5%	Score 582;	DB 16;	Length 306;
Best Local Similarity	44.3%	Pred. NO. 2.5e-53;		
Matches 117; Conservative	40;	Mismatches 67;	Indels 40;	Gaps 5

OY	6	YEQHSJNEIYSWEEFTIERPDMLTIRHISSEFEKPLVLYLKVSGEQGTAKNIMWDG	65
Dd	6	yekynwetteleawckqytsenpdlsrtatagttlgnmlyllkx-gkpgpnkpaimdcg	64
OY	66	IHAEMISPAFLMFI-----GH-----NRMR	88
Dd	65	fhaewshaficqwfivreaulltygeeshmetflnkldfyvlpvlnidgylytclknmrwr	122
OY	89	KNRSFYNANNCICIGDLNRNFASKRMCEGASSSSCSTFYCGLYTESPEPKAVASFURN	148
Dd	125	ktrsnaagtcctgcgdpnrnf-dagwtcttgastcdpcdeylcgssaaesketraladfrnm	183
OY	149	INOIKAYISMHSYSQHITVFPSYSTRSKSCKHEELSLVAASEAVRAIEKTSKNTRTHHGS	208
Dd	184	lsaikeylthhsyqmillypsydqykpiennaelmlakeavnel-atlygtkyfrytpga	242
OY	209	EPLYLAPGGDDWTYDGIKTSFPT	232
Dd	243	tilypaegsgddwaydgglkysat	266

RESULT 12

ID	standard; Protein; 306 AA
AAU04477	standard; Protein; 306 AA

AC AAU04477;

DT 26-SEP-2001 (first entry)

Porcine carboxypeptidase B (CPB) protein.

KW carboxypeptidase B; CPB; p4g: protease activity; trypsin; chymotrypsin;
KW animal product; purification; contaminant; biopharmaceutical agent;
KW zinc-containing pancreatic exopeptidase; insulin; proteolytic processing
KW protein sequence analysis.

OS Sus scrofa.

PN W0200151624-A2

PD 19-JUL-2001.
yy

12-JAN-2001; 2001WO-US000008

12-JAN-2000; 2000US-0175781

(ELIL) LILLY & CO ELI.

PL Hanquillet JM;
XX

WFL; 2001-442143/41

PT New carboxypeptidase B, which is free of animal products and
PT contaminating enzyme activity, useful for producing biopharmaceutical
PT agents (e.g., insulin) and biologically active polypeptides, as well as
PT in protein sequence analysis -
XX

PS Claim 4; Page 41-42; 42pp; English.
XX

The sequence represents porcine carboxypeptidase B (CPB), which is essentially free of contaminating protease activity (e.g. trypsin or chymotrypsin activity) and free of animal products, bacterial cell components and other contaminants that result from purification processes. CPB is a zinc-containing pancreatic exopeptidase which specifically removes C-terminal arginine, lysine and ornithine from peptides. The CPB of the invention is useful for commercial and research purposes, e.g. in production of biopharmaceutical agents, such as insulin and other biologically active polypeptides, as well as in protein sequence analysis. Compositions containing CPB are useful in preparing proteins which are in need of proteolytic processing in order to form the mature protein.

sq Sequence 306 AA;

Query Match	43.58;	Score 582;	DB 22;	Length 306;
Best Local Similarity	44.38;	Pred. No. 2.5e-53;		
Matches 117; Conservative	40;	Mismatches 67;	Indels 40;	Gaps 5

QY		6	YEOYSHINATYIWMIEFIIERHPDLTITHIGSFEKXPYLVLTKVSGKEQLAKNAIMWDCG	65
Dd		6	yekymwmetlaewtkqyvsaenpdlsitalygtifgmnyllkv-gk9gpnkpafimdcg	64
QY		66	HAREWISPAFLMFI-----GH-----NRMR	88
Dd		65	fnaewelshafocqfveazeavltyegshmfeflnkidfyvlpvlnldgylytwcknmwr	122
QY		89	KRSEFYANHCIGDPLNPNFKSMHCMBEGASSSCSETCGILPSPBEPYKAVASFLRRN	148
Dd		125	ketscnagttclcygcpnrf-dagwcttgsatstpcdecygaaseeetckaladflnn	163
QY		149	INOIAVTISMHSOIHVPYSYTRSKSKCHEELSLVAEAVRAIEKTSKMTPTHTHGS	208
Dd		184	lsslkayltihnsyqmllpypsydyukipennaelnmlakaavkel-attlygkytyvgpa	242
QY		209	EFLYLAPGGDWLYLDGIKYSPF	232
Dd		243	ttiypaagsddawaydgqlkystl	266

RESULT 13

AA075131 standard; Protein; 404 AA.

AC AAR75131;
XY

13-MAR-1996 (first entry)

De Porcine Tyr-His-Met Procarboxypeptidase B.

procarboxypeptidase B; carboxypeptidase B; pcpb;
human serum albumin; premating factor alpha;
procbb.

05 Sus scrofa

PN W09514096-A1
XY

PD 26-MAY-1995

16-NOV-1994; 94WO-US13142.

PR 16-NOV-1993; 9305-0153258.

PA (ELL) LIBBY & CO ELL

XX Fayerman JT, Greenen DP, Hersberger CL, Larson JL;
PI Steiner JL, Zhang H;
XX WPI: 1995-200386/26.
DR N-PSDB: AAO90600.
XX
PT DNA encoding porcine carboxypeptidase B - used for transforming
PT host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS Example 1; Page 16-17; 34pp; English.
XX
CC The porcine carboxypeptidase B coding sequence can be place in a
CC bacterial or pref. Pichia yeast expression vector. The expression
CC vector further comprises the signal peptide of either human serum
CC albumin (designated PLCD23 - NRRL B-21029); premating factor alpha
CC (designated PFJ489 - NRRL B-21028); mating factor alpha (designated
CC PFJ474 - NRRL B-21032) or the porcine proBb signal peptide.
CC (designated PL6D27 - NRRL B-21027). The method can be used for
CC producing large amounts of porcine carboxypeptidase B and when
CC produced in Pichia yeast, the protein does not need solubilisation
CC or folding. The produced enzyme is then used for pref. cleaving
CC basic residues from the carboxy terminus of proteins.
XX
SQ Sequence 404 AA:

Query Match 43.5%; Score 582; DB 16; Length 404;
Best Local Similarity 44.3%; Pred. No. 3,8e-53;
Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;

QY 6 YEQHSLEIYSWIEFIERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDG 65
Db 104 yekynwctleawtqytsenpdlistraigtigniylikv-gkpgynkpaifmdcg 162
QY 66 IHAREWISPAFCMLFT-----GH-----NRMR 88
Db 163 fharewishaicqwfvrcaavltgyeshtmfinkldfyvlpvlnidgyitytwcknrmr 222
QY 89 KNRSFYANNHCIGTDLNNRNFASKHWCCEGASSSSCSEFYCGLYPSEPEVKAASFLRRN 148
Db 223 ktrstnagttcigtcpnrf-dagwcttgastdpdcetycsaaseketkaladlrnn 281
QY 149 INQIAVISMYSQHIYFPYSYTRSKSKDHELSVASEAVRAIEKSKNTRYTHGCS 208
Db 282 issikayitlthysgmllpyysydykklpenaealnlaakaavkel-atlygtktyygsa 340
QY 209 ETLVLAPEGGDWIDLGIKYSFT 232
Db 341 tllypaagsgddwaydgikysft 364

RESULT 14
AAW06174
ID AAW06174 standard; Protein: 329 AA.
XX
AC AAW06174;
XX
DT 17-FEB-1997 (first entry)
XX
DE Mature HCPB with PeLB leader sequence.
XX
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPPT;
KW mustard-ribonuclease; antibody directed enzyme produg therapy;
KW anti-neoplastic; produg; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW F(ab')2; PeLB; leader; human carboxypeptidase B.
XX
XX Synthetic.
XX
OS
XX
XX Key location/Qualifiers
XX 1..22
XX Peptide /label= PeLB_leader_sequence
XX
XX

FT Protein 23..329
FT /label= mature_HCPB
XX
XX W09620011-A1.
XX
XX 04-JUL-1996.
XX
XX 21-DEC-1995; 95W0-GB02991.
XX
XX 16-AUG-1995; 95GB-0016810.
XX
XX 23-DEC-1994; 94GB-0026192.
XX
XX (ZENEC) ZENECA LTD.
XX
XX Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
XX Henham JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
XX Tarragona-Piol A, Taylorson CJ;
XX WPI: 1996-321650/32.
XX N-PSDB: AAT42500.
XX
XX Two component system for anti-tumour therapy - comprising targeting
XX moiety linked to mutated enzyme which can transform an
XX anti-neoplastic produg
XX
XX
XX Reference Example 16; Page 137-139; 182pp; English.
XX
XX A two-component system for anti-tumour therapy comprises a targeting
XX moiety linked to a mutated enzyme which can transform an anti-neoplastic
XX produg. The system is based on antibody directed enzyme produg therapy
XX (ADPPT) using non-naturally occurring mutant forms of host enzymes,
XX pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
XX CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
XX be used. The mature native HCPB and PeLB leader sequence coding sequences
XX were cloned (using primers AAT42498-99) from pIC11712 (see AAT42497) for
XX expression in E. coli. The present sequence is that of the mature HCPB
XX and PeLB leader sequence.
SQ Sequence 329 AA:

Query Match 42.4%; Score 567; DB 17; Length 329;
Best Local Similarity 41.3%; Pred. No. 1,1e-51;
Matches 111; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

QY 1 ASASYEQHSLEIYSWIEFIERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAI 60
Db 23 atghsyekynkwetleawqgvatenpalrsrvigttegraiylikv-gkagqnpkal 81
QY 61 WIDCGIHAREWISPAFCMLFT-----GH----- 83
Db 82 fmdcgfharewispafcqwfvreaavrtgyreigtvtellkldfyvlpvlnidgyitytwck 141
QY 84 NRMWRKNSRFYANNHCIGTDLNNRNFASKHWCCEGASSSSCSEFYCGLYPSEPEVKAAS 143
Db 142 strfwktrstlthyscigtcpnrf-dagwcttgastdpdcetycsaaseketkalad 200
QY 144 FLRRNIQIAVISMYSQHIYFPYSYTRSKSKDHELSVASEAVRAIEKSKNTRYTH 203
Db 201 flrnkissikayitlthysgmllpyysydykklpenaealnlaakaavkel-aslhgtkyt 259
QY 204 HGHGSETLVLAPEGGDWIDLGIKYSFT 232
Db 260 yspgattltpaagsgddwaydgikysft 288

RESULT 15
AAW06173
ID AAW06173 standard; Protein: 349 AA.
XX
AC AAW06173;
XX
DT 12-FEB-1997 (first entry)
XX

XX	mature HCPB-(His)6-c-myc with PelB leader sequence.
DE	
XX	ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPPT;
KW	mustard-ribonuclease; antibody directed enzyme prodng therapy;
KW	anti-neoplastic; prodng; reverse polarity; ion pair interaction;
KM	reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
Fd; F(ab') ₂ ; PelB; leader; human carboxypeptidase B.	
XX	
OS	Synthetic.
XX	
XX	Key
FT	Location/Qualifiers
FT	1..22
FT	/label= PelB_leader_sequence
FT	23..329
FT	/label= mature_HCPB
FT	330..335
FT	/note= "His6 tag"
FT	336..338
FT	/note= "tripeptide linker"
FT	339..348
FT	/label= c-myc-epitope
FT	/note= "recognised by antibody 9E10"
XX	
XX	W09620011-A1.
PN	
XX	
PD	04-JUL-1996.
XX	
XX	21-DEC-1995; 95WO-GB02991.
XX	
PR	16-AUG-1995; 95GB-0016810.
PR	23-DEC-1994; 94GB-0026192.
XX	
PA	(ZENEC) ZENECA LTD.
PI	Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
PI	Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
PI	Tarragona-Floj A, Taylorson CJ;
XX	
DR	WPI: 1996-321650/32.
DR	N-PSDB; AAT42497.
XX	
PT	Two component system for anti-tumour therapy - comprising targeting
PT	molety linked to mutated enzyme which can transform an
PT	anti:neoplastic prodng
XX	
PS	Reference Example 15; Page 134-136; 182pp; English.
XX	
CC	A two-component system for anti-tumour therapy comprises a targeting
CC	molety linked to a mutated enzyme which can transform an anti-neoplastic
CC	prodng. The system is based on antibody directed enzyme prodng therapy
CC	(ADPPT) using non-naturally occurring mutant forms of host enzymes,
CC	pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC	Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC	be used. The mature native HCPB coding sequence (see AAT42494) was cloned
CC	into the secretion vector pIC1266 (which contains a PelB translation
CC	leader sequence) to help purification and detection of the expression
CC	construct a C-terminal peptide tag, (His)6-c-myc was added to the enzyme.
CC	The tag consists of 6 histidyls, a tri-peptide linker (EPB) and a
CC	peptide sequence (EGKLISEEDL) from c-myc which is recognised by antibody
CC	9E10. The C-terminal is completed by addition of an asparagine. The
CC	present sequence is that of the mature HCPB-(His)6-c-myc protein with
CC	PelB leader sequence.
XX	
XX	Sequence 349 AA;

```

Db      23 atgnsyekynkwetleawtqvatempallsrsvigtgfegralytlkv-gkagqnkpa1 81
Oy      61 WIDCGIAREKEMISAFCLMFL-----GH----- 83
Db      82 fmdcgfharewleispafcgwiftravrtfygreigtelldkldfyjvlnldgytwtck 141
Oy      84 NRMWRKNRSPFANNHCICTDNRNRFASKHMCBEQASSSCSEFYCGLPSEPEYKAVAS 143
Db      142 srfrwktstthtsgscigtjpnrrnf-dagwkeiysarjnpcedocygpaaseeketralad 200
Oy      144 FLRRNINQIKAYISMHSTSGHIVPPYITSTNSKSDHEELSLVASEAVRALEKTSKNTRYT 203
Db      201 firnklsiskaylthtssysgmmylpyrsayxklgenaealnalakatykel-aehlgstkyt 259
Oy      204 HGHGSEITLYLAPGGDDMTYLYGKIKSYPT 232
Db      260 yypgacttlypaaagsddawaydqgrtstft 288

```

Search completed: September 18, 2002, 16:45:13
Job time: 380 sec

	Query Match	42.4%	Score 567;	DB 17;	Length 349;
	Best Local Similarity	41.3%;	Pred. NO. 1.2e-51;		
	Matches 111; Conservative	46;	Mismatches 72;	Indels 40;	Gaps 5.
QY	1 ASASYEEQHSHNELIYSWIEFTTERHPDMLTTHIHGSSFEKPIVLTKXSGKEQTAKNAI	60			
	: :	: : : :	:	:	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: September 18, 2002, 16:49:06 ; Search time 55.22 Seconds

(without alignments)
770.676 Million cell updates/sec

Title: US-09-980-881-4

Perfect score: 1338
Sequence: 1 ASASYEQYHSLNEITSMIE.....IKYFPTSNPVEKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

SPTREMBL_19:**
1: sp_archaea:**
2: sp_bacteria:**
3: sp_fungi:**
4: sp_human:**
5: sp_invertebrate:**
6: sp_mammal:**
7: sp_mmc:**
8: sp_organelle:**
9: sp_phage:**
10: sp_plant:**
11: sp_rodent:**
12: sp_virus:**
13: sp_vertebrate:**
14: sp_unclassified:**
15: sp_virus:**
16: sp_bacteriaph:**
17: sp_archaeap:**

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	1327	99.2	360	4	Q9P2Y6	Q9P2Y6 homo sapien
2	1239.5	92.6	423	4	Q15114	Q15114 homo sapien
3	1234.5	92.3	423	4	Q961Y4	Q961Y4 homo sapien
4	1105.5	82.6	422	11	Q9UJH6	Q9UJH6 mus musculu
5	1095.5	81.9	422	11	Q9EYV9	Q9EYV9 ratu musculu
6	1095.5	81.9	422	11	Q9QZFO	Q9QZFO mus musculu
7	612.5	45.8	416	13	Q9PUF2	Q9PUF2 bothrops ja
8	582	43.5	416	6	Q9XSP3	Q9XSP3 sus scrofa
9	567	42.4	417	4	Q96BQ8	Q96BQ8 homo sapien
10	557	41.6	412	11	P97397	P97397 ratu musculu
11	552	41.3	417	4	Q96E94	Q96E94 homo sapien
12	537	40.1	279	11	Q9CVD1	Q9CVD1 mus musculu
13	488.5	35.5	419	6	Q9TV85	Q9TV85 sus scrofa
14	477.5	35.7	247	4	Q96K29	Q96K29 homo sapien
15	465.5	34.8	419	4	Q9BS67	Q9BS67 homo sapien
16	459	34.3	417	4	Q96QNS	Q96QNS homo sapien

17	459	34.3	417	4	Q96A12	Q96A12 homo sapien
18	456	34.1	198	4	Q9NT18	Q9NT18 homo sapien
19	426	31.8	1192	5	Q9W475	Q9W475 drosophila
20	424	31.7	424	5	Q9YL87	Q9YL87 drosophila
21	419	31.3	422	5	Q9YL82	Q9YL82 drosophila
22	419	31.3	440	5	Q9W478	Q9W478 drosophila
23	403	30.1	323	5	Q19121	Q19121 caenorhabd
24	396.5	29.6	584	5	Q9RZM1	Q9RZM1 caenorhabd
25	385.5	28.8	453	5	Q9V342	Q9V342 drosophila
26	382.5	28.6	424	5	Q9VS66	Q9VS66 drosophila
27	380	28.4	430	5	Q9YL86	Q9YL86 drosophila
28	377.5	28.2	423	5	Q61532	Q61532 drosophila
29	375	28.0	427	5	Q9U9K2	Q9U9K2 aedes aegy
30	370.5	27.7	419	5	Q9VLZ1	Q9VLZ1 drosophila
31	366.5	27.4	455	5	Q9TZC6	Q9TZC6 caenorhabd
32	357.5	26.7	400	5	Q9VSM7	Q9VSM7 drosophila
33	355	26.5	467	5	Q9VCM8	Q9VCM8 drosophila
34	354	26.5	581	5	Q23318	Q23318 caenorhabd
35	349	26.1	415	5	Q9VS64	Q9VS64 drosophila
36	344.5	25.7	418	5	Q9VS68	Q9VS68 drosophila
37	344.5	25.7	418	5	Q961J8	Q961J8 drosophila
38	344.5	25.7	540	5	Q9RZH2	Q9RZH2 caenorhabd
39	342	25.6	666	5	Q9XU75	Q9XU75 caenorhabd
40	340	25.4	1430	5	Q9XR86	Q9XR86 drosophila
41	336	25.1	545	5	Q9N3S6	Q9N3S6 caenorhabd
42	331.5	24.8	161	13	Q90W12	Q90W12 brachydanio
43	325.5	24.3	312	5	Q9VRZ3	Q9VRZ3 drosophila
44	319.5	23.9	173	11	Q9D961	Q9D961 mus musculu
45	302	22.6	344	5	Q9VS65	Q9VS65 drosophila

ALIGNMENTS

RESULT 1
ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B-LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL Submitted (MAR-1998) to the EMBL/Genbank/DBS databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=92042093; PubMed=1939207;
RX Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL: AB011969; BAA90475.1; -;
DR HSSP: P00730; 2CTC.
DR MEROPS: M14.009; -;
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF002244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept; 2.
DR PRINTS: PR00765; CRBOXYPTASEA.
SQ SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

Query Match 99.2%; Score 1327; DB 4; Length 360;
Best Local Similarity 99.2%; Pred. No. 6,9e-121;
Matches 244; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 ASASYEQYHSLNFIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKVGKQOTAKNAI 60
DB 115 ASASYEQYHSLNFIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKVGKQOTAKNAI 174
QY 61 WIDCGIHAREWISPAFCFLMFIQHNHMRKNSFYANNHCIGDNLNRFASKWCEGAS 120
DB 175 WIDCGIHAREWISPAFCFLMFIQHNHMRKNSFYANNHCIGDNLNRFASKWCEGAS 234
QY 121 SSCSEYTCGLYPESEPEYKAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 180
DB 235 SSCSEYTCGLYPESEPEYKAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 294
QY 181 ELISVAEAVRAIEKTSKTRTHGHGSETTYLAPGGDDWIYDLGIKYSTSNPVEKL 240
DB 295 ELISVAEAVRAIEKTSKTRTHGHGSETTYLAPGGDDWIYDLGIKYSTSNPVEKL 354
QY 241 LPLSLK 246
DB 355 LPLSLK 360

```

```

RESULT 2
Q15114 PRELIMINARY; PRT; 423 AA.
AC Q15114:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PCPB PROTEIN.
GN PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER.
RX MEDLINE-92042093; PubMed-1939207;
RA Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Dayana D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
RL J. Biol. Chem. 266:21833-21838 (1991).
DR EMBL; M75106; AAA60042.1; -
DR HSSP; P00730; 1PT.
DR MEROPS; M14.009; -
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
FT CHAIN 23
SQ SEQUENCE 423 AA; 48442 MW; 9B383272FE6E79F4 CRC64;

```

```

Query Match 92.6%; Score 1239.5; DB 4; Length 423;
Best Local Similarity 86.2%; Pred. No. 2.8e-112;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```

```

QY 1 ASASYEQYHSLNFIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKVGKQOTAKNAI 60
DB 115 ASASYEQYHSLNFIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKVGKQOTAKNAI 174
QY 61 WIDCGIHAREWISPAFCFLMFIQHNHMRKNSFYANNHCIGDNLNRFASKWCEGAS 120
DB 175 WIDCGIHAREWISPAFCFLMFIQHNHMRKNSFYANNHCIGDNLNRFASKWCEGAS 234
QY 121 SSCSEYTCGLYPESEPEYKAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 180
DB 235 SSCSEYTCGLYPESEPEYKAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 294
QY 181 ELISVAEAVRAIEKTSKTRTHGHGSETTYLAPGGDDWIYDLGIKYSTSNPVEKL 240
DB 295 ELISVAEAVRAIEKTSKTRTHGHGSETTYLAPGGDDWIYDLGIKYSTSNPVEKL 354
QY 241 LPLSLK 246
DB 355 LPLSLK 360

```

```

DB 295 FLRRNINQIKAYISMHSYQHIYFPYSYRSKDHLEISLVAEAVRAIEKTSKTRTYT 354
QY 204 HGHGSETTYLAPGGDDWIYDLGIKYST 232
DB 355 HGHGSETTYLAPGGDDWIYDLGIKYST 383

```

```

RESULT 3
Q0961Y4 PRELIMINARY; PRT; 423 AA.
AC Q0961Y4:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B2 (PLASMA).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007057; AA07057.1; -
SQ SEQUENCE 423 AA; 48412 MW; 9B383E03F6EE7CF5 CRC64;

```

```

Query Match 92.3%; Score 1234.5; DB 4; Length 423;
Best Local Similarity 85.9%; Pred. No. 8.5e-112;
Matches 231; Conservative 0; Mismatches 1; Indels 37; Gaps 1;

```

```

QY 1 ASASYEQYHSLNFIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKVGKQOTAKNAI 60
DB 115 ASASYEQYHSLNFIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKVGKQOTAKNAI 174
QY 61 WIDCGIHAREWISPAFCFLMFIQHNHMRKNSFYANNHCIGDNLNRFASKWCEGAS 120
DB 175 WIDCGIHAREWISPAFCFLMFIQHNHMRKNSFYANNHCIGDNLNRFASKWCEGAS 234
QY 121 SSCSEYTCGLYPESEPEYKAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 180
DB 235 SSCSEYTCGLYPESEPEYKAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 294
QY 181 ELISVAEAVRAIEKTSKTRTHGHGSETTYLAPGGDDWIYDLGIKYSTSNPVEKL 240
DB 295 ELISVAEAVRAIEKTSKTRTHGHGSETTYLAPGGDDWIYDLGIKYSTSNPVEKL 354
QY 204 HGHGSETTYLAPGGDDWIYDLGIKYST 232
DB 355 HGHGSETTYLAPGGDDWIYDLGIKYST 383

```

```

RESULT 4
Q09JH6 PRELIMINARY; PRT; 422 AA.
AC Q09JH6:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE R (THROMBIN-ACTIVATABLE FIBRINOLYSIS INHIBITOR)
DE (1110032P04RIK PROTEIN).
GN CPB2 OR TAFI OR 1110032P04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20341711; PubMed-10876383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA Campbell W., Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,

```


RT Whereas carboxypeptidase N is not.;

RL J. Immunol. 165:1053-1058(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE-20201996; PubMed-10739389;

RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,

RA van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.,

RT "Characterization of mouse thrombin-activatable fibrinolysis

RL inhibitor.;"

RL Thromb. Haemost. 83:297-303(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-C57BL/6J; TISSUE-EMBRYO;

RX MEDLINE-21085660; PubMed-11217851;

RA Kawai J., Shitagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,

RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,

RA Felschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,

RA Schiml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Noridone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Williams L.,

RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse CDNA collection.;"

RL Nature 409:685-690(2001).

DR EMBL; AB021968; BAB3402.1; -

DR EMBL; AF164524; AAF62385.1; -

DR EMBL; AK004045; BAB2141.1; -

DR HSSP; P00730; 2CTC.

DR MEROPS; M14.009; -

DR MGD; MGI:1891837; Cpb2.

DR InterPro; IPR003146; Propep_M14.

DR InterPro; IPR000834; Zn_carboxypept.

DR Pfam; PF00244; Propep_M14; 1.

DR Pfam; PF00246; Zn_carboxypept; 1.

DR PRINTS; PR00765; CRBOXYPTASEA.

DR SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

Query Match 82.6%; Score 1105.5; DB 11; Length 422;

Best Local Similarity 77.0%; Pred. No. 3.1e-99;

Matches 207; Conservative 10; Mismatches 15; Indels 37; Gaps 1;

QY 1 ASASYEQYHSLNEIYSWIFETTERHPDMLTKIHGSEFEKYPYLVKSGKEQTKAKAI 60

DB 114 ASASYEQYHSLNEIYSWIFETTERHPDMLTKIHGSEFEKYPYLVKSGKEQTKAKAI 173

QY 61 WIDCGIHAREMISPAFCLEWFGH----- 83

DB 174 WIDCGIHAREMISPAFCLEWFGH----- 233

QY 84 NRMWRNRNRFYANNHCIGTDLNRNFASKHWCCEGASSSCSEYTCGLYPESEPEVKA 143

DB 234 NRMWRNRNRFYANNHCIGTDLNRNFASKHWCCEGASSSCSEYTCGLYPESEPEVKA 293

QY 144 FLRRNINQIKAYISMHSYSHIYFPPSYTRSKSKDHEELSLVASEAVRAIEKTSKNT 203

DB 294 FLRRNINQIKAYISMHSYSHIYFPPSYTRSKSKDHEELSLVASEAVRAIEKTSKNT 353

QY 204 HGHSEETLYLAPGGDDWIDYLGIKYSFT 232

DB 354 HGHSEETLYLAPGGDDWIDYLGIKYSFT 382

RESULT 5

ID 09EQV9 PRELIMINARY; PRT; 422 AA.

AC 09EQV9;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, last annotation update).

DE PRE-PROCARBOXYPEPTIDASE R.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID-10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-20471387; PubMed-11021404;

RA Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,

RA Okada N., Okada H.;

RT "Molecular cloning and partial characterization of rat

RL procarboxypeptidase R and carboxypeptidase N.;"

RL Microbiol. Immunol. 44:719-728(2000).

DR HSSP; P00730; 2CTC.

DR InterPro; IPR003146; Propep_M14.

DR InterPro; IPR000834; Zn_carboxypept.

DR Pfam; PF00244; Propep_M14; 1.

DR Pfam; PF00246; Zn_carboxypept; 1.

DR PRINTS; PR00765; CRBOXYPTASEA.

KW Carboxypeptidase.

SO SEQUENCE 422 AA; 48826 MW; FFFD32A51A9366C8 CRC64;

Query Match 81.9%; Score 1095.5; DB 11; Length 422;

Best Local Similarity 75.8%; Pred. No. 2.9e-98;

Matches 204; Conservative 11; Mismatches 17; Indels 37; Gaps 1;

QY 1 ASASYEQYHSLNEIYSWIFETTERHPDMLTKIHGSEFEKYPYLVKSGKEQTKAKAI 60

DB 114 ASASYEQYHSLNEIYSWIFETTERHPDMLTKIHGSEFEKYPYLVKSGKEQTKAKAI 173

QY 61 WIDCGIHAREMISPAFCLEWFGH----- 83

DB 174 WIDCGIHAREMISPAFCLEWFGH----- 233

QY 84 NRMWRNRNRFYANNHCIGTDLNRNFASKHWCCEGASSSCSEYTCGLYPESEPEVKA 143

DB 234 NRMWRNRNRFYANNHCIGTDLNRNFASKHWCCEGASSSCSEYTCGLYPESEPEVKA 293

QY 144 FLRRNINQIKAYISMHSYSHIYFPPSYTRSKSKDHEELSLVASEAVRAIEKTSKNT 203

DB 294 FLRRNINQIKAYISMHSYSHIYFPPSYTRSKSKDHEELSLVASEAVRAIEKTSKNT 353

QY 204 HGHSEETLYLAPGGDDWIDYLGIKYSFT 232

DB 354 HGHSEETLYLAPGGDDWIDYLGIKYSFT 382

RESULT 6

ID 09QZFO PRELIMINARY; PRT; 422 AA.

AC 09QZFO;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE CARBOXYPEPTIDASE U.

GN Cpb2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID-10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-B6/CRJLJ;

RA He Y.C., Broze G.;

RT "Isolation and characterization of mouse liver carboxypeptidase B
 RT gene." submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL; A181818; AAF00528.1; -
 DR HSSP; P00730; 5CPA.
 DR MEROPS; M14.009; -
 DR MGD; MG1:1891837; Cpb2.
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_carboxypept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_carboxypept; 1.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 SQ SEQUENCE 422 AA; 48832 MW; B53FFB09943954E CRC64;

Query Match 81.9%; Score 1095.5; DB 11; Length 422;
 Best Local Similarity 76.2%; Pred. No. 2.9e-98;
 Matches 205; Conservative 11; Mismatches 16; Indels 37; Gaps 1;

OY 1 SASAYEYOHSHINETYSWIEFTEHRPDMLTIKHIGSSFEKPYLVKSGKEQTAKNAI 60
 DB 114 SASAYEYOHSHINETYSWIEFTEHQPDMLQKITYIGSSFEKPYLVKSGKEQRIKNAI 173
 OY 61 WIDCGIHAREWISPAFCIMFI----- 83
 DB 174 WIDCGIHAREWISPAFCIMFIQYQFHCKENLYRLRHDFITMPVANDGYDTWK 233
 OY 84 NMMWRKNSFYANNHCTGDLNRNFAKHWCEGASSSCSEYCYGLYPESEPEYKAVAS 143
 DB 234 NMMWRKNSFYANNHCTGDLNRNFAKHWCEGASSSCSEYCYGLYPESEPEYKAVAD 293
 OY 144 FLRRNINQIKAYISMSYSOHIVPEYSTRSKSDHEELSLVASEAVAIETSKNRYT 203
 DB 234 FLRRNINQIKAYISMSYSOHIVPEYSTRSKSDHEELSLVASEAVAIETSKNRYT 353
 OY 204 HGHGSETLYLAPGGGDDMIYDLGIRYSFT 232
 DB 354 HGHGSETLYLAPGGGDDMIYDLGIRYSFT 382

RESULT 7
 OY 09PUF2 PRELIMINARY; PRT; 416 AA.
 AC 09PUF2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CARBOXYPEPTIDASE HOMOLOG.
 OS Bothrops jararaca (Jararaca).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
 OC Viperidae; Crotalinae; Bothrops.
 OX NCBI_TaxID=8724;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
 RT "Screening of Bothrops jararaca pancreas cDNA library."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; A190274; AAF01344.1; -
 DR HSSP; P09955; 1NSA.
 DR MEROPS; M14.003; -
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_carboxypept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_carboxypept; 1.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 SQ SEQUENCE 416 AA; 47723 MW; 4F999854DD72B7A7 CRC64;

Query Match 45.8%; Score 612.5; DB 13; Length 416;
 Best Local Similarity 45.0%; Pred. No. 2.9e-51;

Matches 121; Conservative 44; Mismatches 63; Indels 41; Gaps 5;
 OY 2 SASY-YEOHSHINETYSWIEFTEHRPDMLTIKHIGSSFEKPYLVKSGKEQTAKNAI 60
 DB 111 TAGYTERKINSWEKIDAMTADIANENPSVSLQIGTFEGRPMLKV-GKPGVKNKAI 169
 OY 61 WIDCGIHAREWISPAFCIMFI-----GH 83
 DB 170 FIDCGFHAREWISPAFCIMFVREAVRTYQKETIMQTLNKIDFYILPLVINDGYVYSNK 229
 OY 84 NMMWRKNSFYANNHCTGDLNRNFAKHWCEGASSSCSEYCYGLYPESEPEYKAVAS 143
 DB 230 NMMWRKNSFYANNHCTGDLNRNFAKHWCEGASSSCSEYCYGLYPESEPEYKAVAD 288
 OY 144 FLRRNINQIKAYISMSYSOHIVPEYSTRSKSDHEELSLVASEAVAIETSKNRYT 203
 DB 269 FLRRNINQIKAYISMSYSOHIVPEYSTRSKSDHEELSLVASEAVAIETSKNRYT 347
 OY 204 HGHGSETLYLAPGGGDDMIYDLGIRYSFT 232
 DB 348 YGPGCATIYLPAAAGSGDDMIYDLGIRYSFT 376

RESULT 8
 OY 09XSP3 PRELIMINARY; PRT; 416 AA.
 AC 09XSP3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PROCARBOXYPEPTIDASE B PRECURSOR (EC 3.4.17.2).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PANCREAS;
 RX MEDLINE=93121926; PubMed=10391940;
 RA Ventura S., Villegas V., Sturner J., Larson J., Vendrell J.,
 RA Herschberger C., Aviles F.;
 RT "Mapping the pro-region of carboxypeptidase B by protein engineering.
 RT Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
 RL J. Biol. Chem. 274:11925-11933(1999).
 DR EMBL; AJ13775; CAB46991.1; -
 DR HSSP; P09955; 1PBA.
 DR MEROPS; M14.003; -
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_carboxypept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_carboxypept; 1.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 KM Signal; Carboxypeptidase; Hydrolase.
 FT STGNL 1 15 POTENTIAL.
 FT CHAIN 16 416 PROCARBOXYPEPTIDASE B.
 SQ SEQUENCE 416 AA; 47381 MW; 5FA7FCEB7B45AE6A CRC64;

Query Match 43.5%; Score 582; DB 6; Length 416;
 Best Local Similarity 44.3%; Pred. No. 2.7e-48;
 Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;
 OY 6 YEOYHSHINETYSWIEFTEHRPDMLTIKHIGSSFEKPYLVKSGKEQTAKNAIIDDG 65
 DB 116 YEKYHSHINETYSWIEFTEHRPDMLTIKHIGSSFEKPYLVKSGKEQTAKNAIIDDG 174
 OY 66 IHAREWISPAFCIMFI-----GH-----NRMR 88
 DB 175 IHAREWISPAFCIMFI-----GH-----NRMR 234
 OY 89 KNSFYANNHCTGDLNRNFAKHWCEGASSSCSEYCYGLYPESEPEYKAVASFLRRN 148

DB 235 KRSSTNAGTCTGTDPNRNF-DAGMCTGASTDPDDETCGSAASEKTKALADPFRNN 293
 OY 149 INQIKAYISMHSYQSHIVFPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
 DB 294 ISSIKAYLTHSYQSHIVFPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 352
 OY 209 ETLTYLAPGGDDMIYDLGIKYSFT 232
 DB 353 TTYPAAGSDDMAYDQGIKYSFT 376

RESULT 9
 ID 096B08 PRELIMINARY: PRT: 417 AA.
 AC 096B08:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYDROTHERMAL 47.4 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ADRENAL CARCINOMA;
 RA Strausberg R.;
 DR Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC015338; AAH15338.1;
 KW Hypothetical protein.
 SQ SEQUENCE 417 AA; 47367 MW; EBBB98B27F5D5AF9 CRC64;

Query Match 42.4%; Score 567; DB 4; Length 417;
 Best Local Similarity 41.3%; Pred. No. 7.7e-47;
 Matches 113; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

OY 1 ASASYEYHSLNELYSWIEFTTERHPDMLTKIHGSEFEKPYLYLVKSGEQAKNAI 60
 DB 111 ATGHSYKKNKMETEAMTQVATENPALISRSVIGTPEGRAIYLKV-GRAGQNKPAI 169
 OY 61 WTDCGIHAREWISPAFCIMFI-----GH----- 83
 DB 170 FMDGCFHAREWISPAFCIMFI-----GH----- 229
 OY 84 NEMWKRNSFYANNHCIGTDLNRFASKHWCCEGASSSCSEYCGLYPESEPEVKAVAS 143
 DB 230 SFPMKTSRTHGSSCIGTDPNRNF-DAGMCEIGASRNCDDETCGPAESEKTKALAD 288
 OY 144 FLRRINDQIKAYISMHSYQSHIVFPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYT 203
 DB 289 FIRNKLSSIKAYLTHSYQSHIVFPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYT 347
 OY 204 HGHSEETLYLAPGGDDMIYDLGIKYSFT 232
 DB 348 YGPATTYTPAAGSDDMAYDQGIKYSFT 376

RESULT 10
 ID 097597 PRELIMINARY: PRT: 412 AA.
 AC 097597:
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MAST CELL CARBOXYPEPTIDASE A PRECURSOR (FRAGMENT).
 GN R-CPA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE DAWLEY;
 RX MEDLINE-97149430; PubMed-8996238;
 RA Lutzelschwab C., Pfeiler G., Aveskog M., Hellman L.;
 RT "Secretory granule proteases in rat mast cells. Cloning of 10
 RT different serine proteases and a carboxypeptidase A from various rat
 RT mast cell populations." (1997).
 RL J. Exp. Med. 185:13-29(1997).
 DR EMBL: U67914; AAB48267.1;
 DR HSSP: P09955; INSA.
 DR MEROPS: M14.010;
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_cathopept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_cathopept; 1.
 DR PRINTS: PR00765; CARBOXYPEPTASE.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Signal; Carboxypeptidase.
 FT NON_TER 1
 FT SIGNAL <1 10 POTENTIAL.
 FT CHAIN 11 412 MAST CELL CARBOXYPEPTIDASE A.
 SQ SEQUENCE 412 AA; 47944 MW; DAB59555FC49137D CRC64;

Query Match 41.6%; Score 557; DB 11; Length 412;
 Best Local Similarity 43.0%; Pred. No. 7.1e-46;
 Matches 113; Conservative 38; Mismatches 72; Indels 40; Gaps 5;

OY 6 YQGYHSLNELYSWIEFTTERHPDMLTKIHGSEFEKPYLYLVKSGEQAKNAI 65
 DB 111 YQGYHSLNELYSWIEFTTERHPDMLTKIHGSEFEKPYLYLVKSGEQAKNAI 169
 OY 66 IHAREWISPAFCIMFI-----GH-----RMNR 88
 DB 170 IHAREWISPAFCIMFI-----GH----- 229
 OY 89 KNSFYANNHCIGTDLNRFASKHWCCEGASSSCSEYCGLYPESEPEVKAVASFLRN 148
 DB 230 KNSFYANNHCIGTDLNRFASKHWCCEGASSSCSEYCGLYPESEPEVKAVASFLRN 288
 OY 149 INQIKAYISMHSYQSHIVFPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
 DB 289 LNSIKAYLTHSYQSHIVFPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 347
 OY 209 ETLTYLAPGGDDMIYDLGIKYSFT 231
 DB 348 STTYKTSGLDMAYDQGIKYSFT 370

RESULT 11
 ID 096E94 PRELIMINARY: PRT: 417 AA.
 AC 096E94:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE SIMILAR TO CARBOXYPEPTIDASE A3 (MAST CELL).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BONE MARROW, AND ACUTE MYELOCYTIC LEUKEMIA;
 RA Strausberg R.;
 DR Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC012613; AAH12613.1;
 KW Carboxypeptidase.
 SQ SEQUENCE 417 AA; 48714 MW; F781EEF23F570E98 CRC64;

Query Match 41.3%; Score 552; DB 4; Length 417;

[illegible]

Query Match	40.1%	Score 537	DB 11	Length 279
09CVD1	12			
09CVD1	PRELIMINARY;	PRT;	279 AA.	
AC 09CVD1:				
DT 01-JUN-2001 (TREMBLrel. 17, Created)				
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE 2210008M23R1K PROTEIN (FRAGMENT) .				
GN 2210008M23R1K.				
OS Mus musculus (Mouse) .				
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX NCBI_TaxId=10090;				
PN 11]				
RP SEQUENCE FROM N.A.				
RC STRAIN=C57BL/6J; TISSUE=STOMACH;				
RX MEDLINE=21085660; PubMed=11217811;				
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,				
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,				
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,				
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA Fletschmann W., Gaasterland T., Gisli C., Kling B., Kochia H.,				
RA Knehl P., Lewis S., Matsuo Y., Nikdel I., Pesole G., Quackenbush J.,				
RA Schirral M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA Sakai K., Orido T., Furuno M., Aono H., Baldirelli R., Barsh G.,				
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA Gutschalk S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,				
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,				
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,				
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,				
RA Hayashizaki Y.;				
RT "functional annotation of a full-length mouse cDNA collection.";				
RL Mature 409.685-69012001).				
DR EMBL; AK008678; BAB25826.1; -.				
DR HSSP; P09955; INSA.				
DR MGD; MGI:1933953; 2210008M23R1K.				
DR InterPro; IPR000834; Zn_carpodpt.				
DR Pfam; PF00246; Zn_carpodpt. 1.				
DR PRINTS; PR00765; CARBOXYPRASE.				
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; UNKNOWN.1.				
FT NON_TER 279				
FT TER 279				
SQ SEQUENCE 279 AA; 31970 MW; 29468FA0B67F0AA CRC64;				

Best Local Similarity 46.0%; Pred. No.3,8e-44; Matches 104; Conservative 31; Mismatches 51; Indels 40; Gaps 4;

```
OY      44 LYLVLKVSKEQCTAKNAIIDCGIHNRREMSIPAFCSMTF----- 81
       :::|||: |||: | | | | | | | | | | | | | | | | | |
Db      1 MYVLKI -GKDPRNKPFAFIIDCGFHAREMISPFECMPFYREAVRTYKQETIMRLLDELDF 59
OY      82 -----GNRRMMKRKNSEFYANNHCITDILNRNFASKHMCDEGASSSCSET 126
       : | | | | | | | | | | | | | | | | | | | | | |
Db      60 YLPLPVYNIDGYVTWAKRMRMKRIKSTTASGCCSFVDNPNRNF-DAGWCVEVGSRSPCSDT 118
OY     127 YCGLPESEPEVKAAVASLELRNRINQIKAYISMHSYQHILVFPSYTSRSKKHEELSLVA 186
       ||| | | | | | | | | | | | | | | | | | | | | |
Db     119 YCGPTPESEKETKALADIRQLWSLKIKAYLVTHVSYSQMLLYPYSDYKLKPNYEELNLV 178
OY     187 SEAVRAILEKTNRPYTHGHSEETLYLPAGGGDDMYLYLGITYYST 232
       : : | | | | | | | | | | | | | | | | | | | |
Db     179 KGAAKEL-STLHGRTYTGPGATITTPAAGSGDDMAIVDGIIYST 223
```

Query Match	36.5%	Score 488.5	DB 6	Length 419
Best Local Similarity	39.0%	Pred. No. 3.4e-39		
Matches 104	Conservative 42	Mismatches 74	Indels 47	Gaps 6
QY 6 YEQHSLSNELYSWLEFETTERHPDMLTKIHSSPEKTYLYLKYSGKQIATKN--AIWI 62				
DB 119 YATVHTLEELIDFMDILVAEHPOLVSKLOIGSSYEGRPYVLKES---TGNRRPAIWI 174				
QY 63 DCGIHAREWISPAFCMIWI-----GHNRR 85				
DB 175 DTGHSRKMVYQASGVWFAFKITIEDYGDPAFTALDNLDFLELVTPNDGPAFTHSENR 234				
QY 86 MWRNRSRFYANNHCIGTDLNRNFASKHMCBEGCASSSCSEITYCGLYPSEPEVKAVASFL 145				
DB 235 MWRTRSRVTSISFCVGDVPRNNW--DAGRGAGAGASSNCPSEITYHGKFPSEVEKVSIDVF 293				
QY 146 RRNIINQIKAYISMSYSCHIVPPYSYTRSKSKDKHBEELVLAASEAVRAIEKTSKNTRYHG 205				

DB 294 NDHGN- IKAFTSHSYSQLMTPYGYKTEPAPADKDELQISKAVAL- TSLYGTFRQY 351
 QY 206 HGSFTLYLAPGGDDMIYDLGIRYSFT 232
 DB 352 SITTTIYQASGTTDWTYNQGIKYSFS 378

RESULT 14
 Q96K29 PRELIMINARY; PRT; 247 AA.

AC 096K29;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PANCREAS;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC015104; AAH15104.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 247 AA; 28181 MW; 914DD8BA162C1082 CRC64;

Query Match 35.7%; Score 477.5; DB 4; Length 247;
 Best Local Similarity 43.8%; Pred. No. 2e-38;
 Matches 91; Conservative 30; Mismatches 48; Indels 39; Gaps 4;

QY 62 IDGIIHAREWISPAFLMFI-----GH-----N 84
 DB 1 MCGFIHAREWISPAFCQWVREAVRTYGRIOYTELLDKIDFVLPVNLNDGIYITWTKS 60
 QY 85 RMKRKRSTYANNHCIGTDLNRNFASKHWCCEGASSSCSETYGLYPESEPEYKAVASF 144
 DB 61 RFWKRKRSTYHTGSSCIGTPNRFN- DAGWCEIGASRNPCDEYCGPAESEKETKALADF 119
 QY 145 LRNNIOIKAYISMHSQHIYFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTH 204
 DB 120 IRKKLSIRAYLTISHISQMLTPYSTAYKLGNNALNALAKATYKEL-ASLHGTYTY 178
 QY 205 HGSFTLYLAPGGDDMIYDLGIRYSFT 232
 DB 179 GPGATTIYPAGSGDDMAYDQGIYSFT 206

RESULT 15
 Q9BS67 PRELIMINARY; PRT; 419 AA.
 AC 09BS67;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CARBOXYPEPTIDASE A1 (PANCREATIC).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PANCREAS;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC005279; AAH05279.1; -.
 DR HSSP; P00730; 1PYT.
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_caropept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_caropept; 1.

DR PRINTS; PR00765; CARBOXYPTASE.
 DR PROSITE; PS00132; CARBOXYPEPT_2N_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_2N_2; 1.
 SQ SEQUENCE 419 AA; 47159 MW; 7855759PAPF982DD CRC64;

Query Match 34.8%; Score 465.5; DB 4; Length 419;
 Best Local Similarity 37.9%; Pred. No. 5.9e-37;
 Matches 100; Conservative 44; Mismatches 79; Indels 41; Gaps 5;

QY 6 YEQHSLNELYSMTEFITERHPDMLTKIHGSSPEKPLVIVKYSKEQAKNAIWDGC 65
 DB 119 YATHTLEETDLDLVAENPHLVSKIQIGNYEGRPYLVKFS-TGSKRPPIWIDTG 177
 QY 66 IHAREWISPAFLMFI-----GHNMNR 88
 DB 178 IHSHEWYQASGVFAKKITQDYGDAFTALDITDIFLEIYTNPGFAFTHSTNMMR 237
 QY 89 KNRSEYANNHCIGTDLNRNFASKHWCCEGASSSCSETYGLYPESEPEYKAVASFRRN 148
 DB 238 KTRSHTAGSLCIGVDPNRNDAGFGL-SGASSNPCSETYGRKFANSEVEKSIYDFVKDH 296
 QY 149 INQIKATISMHSQHIYFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHS 208
 DB 297 GN-IKAFISHSYSQLMTPYGYKTEPEVPDQDELQSKAAVVAL-ASLYGTFRNGSII 354
 QY 209 ETLYLAPGGDDMIYDLGIRYSFT 232
 DB 355 KATYQASGTTDWTYSQGIKYSFT 378

Search completed: September 18, 2002, 16:49:07
 Job time: 299 sec

This Page Blank (u.s.)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:46:03 ; Search time 28.95 Seconds

(without alignments)
207.554 Million cell updates/sec

Title: US-09-980-881-4

Perfect score: 1338

Sequence: 1 ASASYEQYHSINLETYSWIE.....IKTSFMSNPPEVKLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: Issued Patents AA.*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/BACKFILES.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239.5	92.6	423	1	US-07-649-591B-3
2	1239.5	92.6	423	1	US-08-277-540-3
3	1239.5	92.6	423	1	US-08-430-787A-3
4	1239.5	92.6	423	2	US-08-869-057-2
5	610	45.6	307	2	US-08-782-760-6
6	610	45.6	307	5	PCT-US96-00995-6
7	582	43.5	306	1	US-08-696-139-4
8	582	43.5	404	1	US-08-696-139-2
9	567	42.4	415	2	US-08-860-882A-57
10	561	41.9	417	1	US-07-649-591B-7
11	561	41.9	417	1	US-08-277-540-7
12	561	41.9	417	1	US-08-430-787A-7
13	554	41.4	396	1	US-07-649-591B-4
14	554	41.4	396	1	US-08-277-540-4
15	554	41.4	396	1	US-08-430-787A-4
16	552	41.3	417	1	US-07-649-591B-6
17	552	41.3	417	1	US-08-277-540-6
18	552	41.3	417	1	US-08-430-787A-6
19	549	41.0	613	4	US-09-171-945-113
20	549	41.0	716	4	US-09-171-945-125
21	487.5	36.4	419	1	US-07-649-591B-5
22	487.5	36.4	419	1	US-08-277-540-5
23	487.5	36.4	419	1	US-08-430-787A-5
24	483	36.1	417	1	US-07-649-591B-8
25	483	36.1	417	1	US-08-277-540-8
26	483	36.1	417	1	US-08-430-787A-8
27	465.5	34.8	419	4	US-08-640-906-2

28	465.5	34.8	419	4	US-09-395-936-2	Sequence 2, Appl
29	460.5	34.4	419	4	US-08-640-906-17	Sequence 17, Appl
30	460.5	34.4	419	4	US-09-395-936-17	Sequence 17, Appl
31	459	34.3	417	4	US-08-640-906-4	Sequence 4, Appl
32	459	34.3	417	4	US-09-395-936-4	Sequence 4, Appl
33	458	34.2	417	4	US-08-640-906-18	Sequence 18, Appl
34	458	34.2	417	4	US-09-395-936-18	Sequence 18, Appl
35	96.5	7.2	734	4	US-08-706-216-2	Sequence 16, Appl
36	95	7.1	706	1	US-08-484-106-16	Sequence 16, Appl
37	95	7.1	706	1	US-08-111-939-13	Sequence 13, Appl
38	90.5	6.8	434	4	US-08-233-989-7	Sequence 7, Appl
39	90.5	6.8	434	1	US-08-111-939-14	Sequence 14, Appl
40	90.5	6.8	435	1	US-08-233-989-3	Sequence 3, Appl
41	89.5	6.7	476	4	US-09-233-989-6	Sequence 6, Appl
42	89.5	6.7	476	4	US-08-111-939-16	Sequence 16, Appl
43	88.5	6.6	435	1	US-08-452-262-2	Sequence 2, Appl
44	88.5	6.6	435	1	US-08-734-550-2	Sequence 2, Appl
45	88.5	6.6	435	1	US-08-734-550-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: NO. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28, 616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELETYPE: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; US-07-649-591B-3

Query Match 92.6%; Score 1239.5; DB 1; Length 423;
Best Local Similarity 86.2%; Pred. No. 5.5e-127;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 ASASYEQYHSINLETYSWIEFITERHPDMLTKIHGSSFEKXPYLVKYSGEQTKNNAI 60
DB 115 ASASYEQYHSINLETYSWIEFITERHPDMLTKIHGSSFEKXPYLVKYSGEQTKNNAI 174
```

QY 61 WIDCIIHAREWISPAFCIMFIGH----- 83
DB 175 WIDCIIHAREWISPAFCIMFIGHITQFYGIQYTNLLRLVDFYVMPVNVNDGYDSMK 234
QY 84 NRMWRKNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVAS 143
DB 235 NRMWRKNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVAS 294
QY 144 FLRRNIQIKAYISMHSYQHIYFPPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203
DB 295 FLRRNIQIKAYISMHSYQHIYFPPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 354
QY 204 HGHSEETLYLAPGGDDMIYDLGIKYSFT 232
DB 355 HGHSEETLYLAPGGDDMIYDLGIKYSFT 383

RESULT 2

US-08-277-540-3
; Sequence 3, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-277-540-3

Query Match 92.6%; Score 1239.5; DB 1; Length 423;
Best Local Similarity 86.2%; Pred. No. 5.5e-127;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 ASASYEQYHSLNETYSWIEFTTERHPDMLTKIHIGSSFEEKYPLVYLKSGKEQTAKNAI 60
|||||

DB 115 ASASYEQYHSLNETYSWIEFTTERHPDMLTKIHIGSSFEEKYPLVYLKSGKEQTAKNAI 174
QY 61 WIDCIIHAREWISPAFCIMFIGH----- 83
DB 175 WIDCIIHAREWISPAFCIMFIGHITQFYGIQYTNLLRLVDFYVMPVNVNDGYDSMK 234
QY 84 NRMWRKNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVAS 143
DB 235 NRMWRKNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVAS 294
QY 144 FLRRNIQIKAYISMHSYQHIYFPPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203
DB 295 FLRRNIQIKAYISMHSYQHIYFPPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 354
QY 204 HGHSEETLYLAPGGDDMIYDLGIKYSFT 232
DB 355 HGHSEETLYLAPGGDDMIYDLGIKYSFT 383

RESULT 3

US-08-430-787A-3
; Sequence 3, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-430-787A-3

Query Match 92.6%; Score 1239.5; DB 1; Length 423;
Best Local Similarity 86.2%; Pred. No. 5.5e-127;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 ASASYEQYHSLNETYSWIEFTTERHPDMLTKIHIGSSFEEKYPLVYLKSGKEQTAKNAI 60
|||||

Qy	1	ASASYEDQYHSLNLTYSWIEETIRHRHDMLTRIHIGSSFKEYPLVYLVKSGKEBTANAI	60
Db	115	ASASYEDQYHSLNLTYSWIEETIRHRHDMLTRIHIGSSFKEYPLVYLVKSGKEBTANAI	174
Qy	61	WIDCGIHAREWISPAFLMTIGH-----	83
Db	175	WIDCGIHAREWISPAFLMTIGHITTOFYGLIGQYTNLTRLYDFYMPVNVNDGYSMK	234
Qy	84	NRMRKNSPFAANNHCIGTDLRNRFASKHMCDEBASSSCSEYFCGLYPSEPPVKVAVS	143
Db	235	NRMRKNSPFAANNHCIGTDLRNRFASKHMCDEBASSSCSEYFCGLYPSEPPVKVAVS	294
Qy	144	FLRRNINQIKAYISMHSYQHIVEPYSTRSKSDHEELSLVASEAVRALEKTSKNTRYT	203
Db	295	FLRRNINQIKAYISMHSYQHIVEPYSTRSKSDHEELSLVASEAVRALEKTSKNTRYT	354
Qy	204	HGHGSELYLAPGGGDDWITDGLIKYSFT	232
Db	355	HGHGSELYLAPGGGDDWITDGLIKYSFT	383

```

1  RESULT      4
2  US-08-869-057-2
3  : Sequence 2, Application US/08869057
4  : Patent No. 5985562
5  :
6  : GENERAL INFORMATION:
7  : APPLICANT: Morser, Michael J
8  : APPLICANT: Nagashima, Mariko
9  : TITLE OF INVENTION: Method of Detecting Thrombotic Disease
10 : TITLE OF INVENTION: Risk
11 : NUMBER OF SEQUENCES: 6
12 :
13 : CORRESPONDENCE ADDRESS:
14 : ADDRESSEE: Berlex Biosciences Legal Department
15 : STREET: 15049 San Pablo Avenue
16 : CITY: Richmond
17 : STATE: California
18 :
19 : COUNTRY: USA
20 : ZIP: 94804-0099
21 :
22 : COMPUTER READABLE FORM:
23 : MEDIUM TYPE: Floppy disk
24 :
25 : COMPUTER: IBM PC compatible
26 : OPERATING SYSTEM: PC-DOS/MS-DOS
27 : SOFTWARE: Patentin Release #1.0, Version #1.30
28 :
29 : CURRENT APPLICATION DATA:
30 : APPLICATION NUMBER: US/08/869,057
31 : FILING DATE: 03-JUN-1997
32 :
33 : CLASSIFICATION: 435
34 :
35 : ATTORNEY/AGENT INFORMATION:
36 : NAME: Washtlien, Wendy L
37 : REGISTRATION NUMBER: 36,301
38 : REFERENCE/DOCKET NUMBER: 51509AUSW1
39 :
40 : TELECOMMUNICATION INFORMATION:
41 : TELEPHONE: 510-262-5411
42 : TELEFAX: 510-262-7095
43 :
44 : INFORMATION FOR SEQ ID NO: 2:
45 :
46 : SEQUENCE CHARACTERISTICS:
47 : LENGTH: 423 amino acids
48 : TYPE: amino acid
49 : STRANDEDNESS:
50 : TOPOLOGY: linear
51 :
52 : MOLECULE TYPE: protein
53 : ORIGINAL SOURCE:
54 : TISSUE TYPE: Plasma
55 :
56 : FEATURE:
57 : NAME/KEY: Peptide
58 : LOCATION: 23..401
59 :
60 : US-08-869-057-2

```

QY	1	ASASYEDYHSLNLTYSIEETTERHPMLTKHIGSSFEXYPLVYLKVSKEOTANAI	60
Db	115	ASASYEDYHSLNLTYSIEETTERHPMLTKHIGSSFEXYPLVYLKVSKEOTANAI	174
QY	61	WIDCGIHAREMISPAFLCMFIGH-----	83
Db	175	WIDCGIHAREMISPAFLCMFIGHITQFYGIIGYXTNLLRLVDYFVMPVNVYDGYSMK	234
QY	84	NRMWRKNSFPANHHCTGTDRNRPASKHWCBEASSSCSEMYCGILYPSSEPVKVAS	143
Db	235	NRMWRKNSFPANHHCTGTDRNRPASKHWCBEASSSCSEYCGILYPSSEPVKVAS	294
QY	144	FLRRNIQIKAYISMHSYQHIVPEYSYTRSKSDHELSLVASEAVRALEKTSKNTRYT	203
Db	295	FLRRNIQIKAYISMHSYQHIVPEYSYTRSKSDHELSLVASEAVRALEKTSKNTRYT	354
QY	204	HGHGSETIYLAPGGGDDMIYDLGIKISYFT	232
Db	355	HGHGSETIYLAPGGGDDMIYDLGIKISYFT	383

```

1      RESULT      5
2      US-08-782-760-6
3      ; Sequence 6, Application US/08782760
4      ; Patent No. 5948668
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Hartman, Jacob
7      ; APPLICANT: Fulga, Netta
8      ; APPLICANT: Mendelovitch, Simona
9      ; APPLICANT: Gorecki, Marian
10     TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
11     TITLE OF INVENTION: CARBOXYPEPTIDASE B
12     NUMBER OF SEQUENCES: 8
13     CORRESPONDENCE ADDRESS:
14     ADDRESSEE: Cooper & Dunham LLP
15     STREET: 1185 Avenue of the Americas
16     CITY: New York
17     STATE: New York
18     COUNTRY: U.S.A.
19     ZIP: 10036
20     COMPUTER READABLE FORM:
21     MEDIUM TYPE: Floppy disk
22     COMPUTER: IBM PC compatible
23     OPERATING SYSTEM: PC-DOS/MS-DOS
24     SOFTWARE: PatentIn Release #1.0, Version #1.30
25     CURRENT APPLICATION DATA:
26     APPLICATION NUMBER: US/08/782,760
27     FILING DATE: 13-JAN-1997
28     CLASSIFICATION: 435
29     PRIOR APPLICATION DATA:
30     APPLICATION NUMBER: 08/378,233
31     FILING DATE: 25-JAN-1995
32     ATTORNEY/AGENT INFORMATION:
33     NAME: White, John P.
34     REGISTRATION NUMBER: 28,678
35     REFERENCE/DOCKET NUMBER: 0336/43847
36     TELECOMMUNICATION INFORMATION:
37     TELEPHONE: (212) 278-0400
38     TELEFAX: (212) 391-0525
39     INFORMATION FOR SEQ ID NO: 6:
40     SEQUENCE CHARACTERISTICS:
41     LENGTH: 307 amino acids
42     TYPE: amino acid
43     TOPOLOGY: linear
44     MOLECULE TYPE: protein
45     US-08-782-760-6

```


[illegible]

RESULT 8
US-08-696-139-2

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	43.5%;	Score 582;	DB 1;	Length 404;
Best Local Similarity	44.3%;	Pred. No. 3.1e-55;		
Matches 117;	Conservative 40;	Mismatches 67;	Indels 40;	Gaps 5

[illegible][illegible]

RESULT 5
PS-09-860-

```

Sequence 57, Application US/0860882A
Patent No. 5985281
GENERAL INFORMATION:
APPLICANT: TAYLORSON, CHRISTOPHER JOHN
APPLICANT: EGGLETON, HENRIKUS JOHANNES
APPLICANT: TARRAGONA-FIOL, ANTONIO
APPLICANT: BABIN, BRIAN ROBERT
APPLICANT: BOYLE, FRANCIS THOMAS
APPLICANT: HENNAM, JOHN FREDERICK
APPLICANT: BLAKELEY, DAVID CHARLES
APPLICANT: MARSHAM, PETER ROBERT
APPLICANT: HEATON, DAVID WILLIAM
APPLICANT: DAVIES, DAVID HUIW
TITLE OF INVENTION: CHEMICAL COMPOUNDS
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.235
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/0860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 415 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-860-882A-57

```

Query Match	42.48;	Score 567;	DB 2;	Length 415;
Best Local Similarity	41.38;	Pred. No. 1.4e-53;		
Matches 111;	Conservative 46;	Mismatches 72;	Indels 40;	Gaps 5

OY 1 ASASYEQOYHSLNYSWIEITERHPDMLTKIHGSSFEXKPLYLVKSQGEQTAKNAI 60
| : ||: | : : : : |||| |
Db 109 ATGHSEKYNNKWETLEAWTOVATENPLISRSYGITTFEGRAITYLKV-GKAGQNKPAT 167

QY 61 WIDCGIHAREMISPAFCIMFI-----GH----- 83
DB 168 FMDGCGHAHEMISPAFCIMFVREAVRTYGREIOVTELLKLDYVPLVINDIGYITWTX 227
QY 84 NRMWRNRSPYANNHICIGDILNRNFASKHMCCEGASSSCSEFYGLYPESEPVAVAS 143
DB 228 SRWRTRSTRTHGSSCIGDIPNRNF-DAGWCELGASRNCDETYGCPAAESEKETALAD 286
QY 144 FLRRNINOKAVITSMHSYQHIYFVPSYTRSKSKCHEELSLVASEAVRAIEKTSKRTYT 203
DB 287 FIRNKLSSIKAVLTIHSYSQMMIYPYSYKLGENNALAKAVTEL-ASLGTXT 345
QY 204 HCHGSEFLYAPGGDDWYDLGIKYSF 232
DB 346 YGPGATTIYPAGSGDDWYDLGIKYSF 374
RESULT 10
US-07-649-591B-7
Sequence 7, Application US/07649591B
Patent No. 5206161
GENERAL INFORMATION:
APPLICANT: Dennis Drayna and Daniel Eaton
TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/649, 591B
FILING DATE: 19910201
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Haesk, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/266-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
US-07-649-591B-7
Query Match 41.9%; Score 561; DB 1; Length 417;
Best Local Similarity 43.7%; Pred. No. 6.5e-53;
Matches 115; Conservative 37; Mismatches 71; Indels 40; Gaps 5;
QY 6 YEOYHSINETYNIETTERHPMLTKIHIGSFEKYPYLVKSGKEQTAKNAIWDG 65
DB 116 YAKYNDMDKTVSWTEKLEKHPMVSRIKIGSTVEDNPLYVKI-GKKGDERRAIFMDG 174
QY 66 IHAREMISPAFCIMFI-----GH-----RMR 88
DB 175 IHAREMISPAFCIMFVQATKSGKKNKIMTKLLDRANFYLVPEVNDGYTWSWTODRMR 234
QY 89 KNRSPYANNHICIGDILNRNFASKHMCCEGASSSCSEFYGLYPESEPVAVASFLRN 148

DB 235 KNRSPYANNHICIGDILNRNF-DVSWDSSPNINKPCLNRYRGPAPSEKETAIVTISH 293
QY 149 INQIAIYSMSYSHIYFVPSYTRSKSKCHEELSLVASEAVRAIEKTSKRTYTHGHS 208
DB 294 LMSIKAYITFHSYQMMIYPYSYKLGENNALAKAVTEL-ASLGTXT 345
QY 209 EFLYAPGGDDWYDLGIKYSF 231
DB 353 SYITKSSGLDWYDLGIKYSF 375

RESULT 11
US-08-277-540-7
Sequence 7, Application US/08277540
Patent No. 5474901
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277, 540
FILING DATE: 19- JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Haesk, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689DICI1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-277-540-7

Query Match 41.9%; Score 561; DB 1; Length 417;
Best Local Similarity 43.7%; Pred. No. 6.5e-53;
Matches 115; Conservative 37; Mismatches 71; Indels 40; Gaps 5;
QY 6 YEOYHSINETYNIETTERHPMLTKIHIGSFEKYPYLVKSGKEQTAKNAIWDG 65
DB 116 YAKYNDMDKTVSWTEKLEKHPMVSRIKIGSTVEDNPLYVKI-GKKGDERRAIFMDG 174
QY 66 IHAREMISPAFCIMFI-----GH-----RMR 88
DB 175 IHAREMISPAFCIMFVQATKSGKKNKIMTKLLDRANFYLVPEVNDGYTWSWTODRMR 234

[illegible]

RESULT 12
 US-08-430-787A-7
 Sequence 7, Application US/08430787A
 Patent No. 5593674
 GENERAL INFORMATION:
 APPLICANT: Drayna, Dennis T., Eaton, Dan L.
 TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: palin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/430,787A
 FILING DATE: 27-APR-1995
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/277,540
 FILING DATE: 19-JUL-1994
 APPLICATION NUMBER: 08/167727
 FILING DATE: 15-DEC-1993
 PRIOR APPLICATION DATA: 07/959944
 FILING DATE: 14-OCT-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/649591
 FILING DATE: 01-FEB-91
 ATTORNEY/AGENT INFORMATION:
 NAME: Hasak, Janet E.
 REGISTRATION NUMBER: 28,616
 REFERENCE/DOCKET NUMBER: 689D1C1D1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-1896
 TELEFAX: 415/952-9681
 TELEX: 910/371-7168
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 417 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-430-787A-7

Query Match	41.9%	Score 561	DB 1:	Length 417
Best Local Similarity	43.7%	Pred. No. 6	5e-53	
Matches 115	Conservative 37	Mismatches 71	Indels 40	Gaps 5
OY	6	YEOHSLMEIYVMIPEFIERHBDMLTKHIGSFEKYPPIYLYAKSGKEETAAKNAIMIDCG	65	
DB	116	YAKIDMDRIYSMTKEMLKEHPDEWYSRIKIGSTIVEDNPFLYIKLT	174	
OY	66	IHAREWISPAFLWFI-----GHN-----RWWR	88	

Db	175							
		I H A R M I S P A C Q M E V I Y O A T S Y C K N K I M T I L D R M N F Y I L P V E N V D G Y I M S T Q O D M R						234
Qy	89	K N R S E Y A N N H I G I D N L N E N F A S K H M C E G A S S C S E T Y C G L Y P E S E P E V A K A V A F I R N						148
Db	235	K N R S K N O N S T I G D L N R N F - D V S M D S S P N T N K C L P N Y R G P A D E S E K T R A V T N F I R S H						293
Qy	149	I N Q I K A Y I S M H S Y S O H I E F P Y S Y R S K S D H E E L S I V A S E A V R A I E K T S K N T R T H G S						208
Db	294	I N S I A V I T T F H S Y S O M L I P I G Y I F K L P P N H O D L L K A R I T D A L - S T R I E T R I I Y G P I A						352
Qy	209	E T L Y L A P G G G D M I Y D L G I K Y S F						231
Db	353	S T I Y K T S G S S L D W Y I D L G I K T F						375

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	Similarity	41.4%	Score	554	DB	1:	Length	396;
	Best Local Similarity		44.3%	Pred.	No. 3	Se-52:			
	Matches	113; Conservative		Mismatches	63;	Indels	40;	Gaps	5
OY	15	IYSWTEFTERHPDMITKIHIGSSFEKYPLVLYKSGEQTAKMAIWDGIGHAREMISP	74						
Dd	104	IEAWIOGVATNDPDLTOSVICTTEFGRRMYVLKI-GKTRPNKPALFIICGHARFMISD	162						
OY	75	AFCLEPI-----GH-----NRMRKRNSRYANN	97						
Dd	163	AFCQCFAREAVTYNOEHMKQLDELDPYLVPVNIIDGYVTWTWKDRMKRKTSTMAAS	222						
OY	98	HCICTDNLNRFASKHWCCEGAASSSCSEFTYYCGLYPESEPDEVAVASFLRNINIQAYTS	157						

Db 223 SCLGVRNRNF-NAGWCEVGASRSPCSETTTCGPAPESKEKTKALADIFRNNLSTIKATLT 281
QY 158 MHSYSQIIVFPYSTRSKSDHELSLVASAVNAIEKTSKNTRYTHGHGSETLYLAPCG 217
Db 282 IHSYSQMLLYPSYDYKLPENYEELNALVGAANEL-ATLHGTYTYEPGATITYPAAG 340
QY 218 GDDWIYDLGIXYSFT 232
Db 341 SDDWSYDGIKYSFT 355

RESULT 14

US-08-277-540-4
Sequence 4, Application US/08277540
Patent No. 5474901
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA: 07/959944
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELETYPE: 910/371-7168
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-277-540-4

Query Match 41.4%; Score 554; DB 1; Length 396;
Best Local Similarity 44.3%; Pred. No. 3.5e-52;
Matches 113; Conservative 39; Mismatches 63; Indels 40; Gaps 5;
QY 15 IYSWIEFTEHRPDMTKIHIGSFEEKYPLVLYKSGKEQTAKNAIMDCIHRREWISP 74
Db 104 IEAMIQVADNDPLVQSYIGTFEGRNMYLKI-GKTRNKPALIFIDCGFHAREWISP 162
QY 75 AFCLMFI-----GH-----NRMWRKNSFYANN 97
Db 163 AFCLMFI-----GH-----NRMWRKNSFYANN 97
QY 98 HCLGTDLNRFASKHMCCEGASSSCSEYCGLYPESEPEVKAVASFLRNNINIKAYIS 157

Db 223 SCLGVRNRNF-NAGWCEVGASRSPCSETTTCGPAPESKEKTKALADIFRNNLSTIKATLT 281
QY 158 MHSYSQIIVFPYSTRSKSDHELSLVASAVNAIEKTSKNTRYTHGHGSETLYLAPCG 217
Db 282 IHSYSQMLLYPSYDYKLPENYEELNALVGAANEL-ATLHGTYTYEPGATITYPAAG 340
QY 218 GDDWIYDLGIXYSFT 232
Db 341 SDDWSYDGIKYSFT 355

RESULT 15

US-08-430-787A-4
Sequence 4, Application US/08430787A
Patent No. 5593674
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELETYPE: 910/371-7168
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-430-787A-4

Query Match 41.4%; Score 554; DB 1; Length 396;
Best Local Similarity 44.3%; Pred. No. 3.5e-52;
Matches 113; Conservative 39; Mismatches 63; Indels 40; Gaps 5;
QY 15 IYSWIEFTEHRPDMTKIHIGSFEEKYPLVLYKSGKEQTAKNAIMDCIHRREWISP 74
Db 104 IEAMIQVADNDPLVQSYIGTFEGRNMYLKI-GKTRNKPALIFIDCGFHAREWISP 162
QY 75 AFCLMFI-----GH-----NRMWRKNSFYANN 97

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:44:08 ; Search time 55.22 Seconds

(without alignments)
1127.819 Million cell updates/sec

Title: us-09-980-881-2

Perfect score: 1911

Sequence: 1 MKLCSLAVLPVIFVFCQHV.....IKYSFTSNPVEKLPLSLK 360

Scoring table: BLOSUM62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1884	98.6	360	4	Q9P2Y6
2	1799.5	94.2	423	4	Q15114
3	1794.5	93.9	423	4	Q961Y4
4	1528	80.0	422	11	Q93H66
5	1522	79.6	422	11	Q9QZFO
6	1499	78.4	422	11	Q9EOV9
7	1029	53.8	198	4	Q9NTI8
8	700.5	36.7	416	13	Q9PUF2
9	646	33.8	416	6	Q9XSP3
10	641	33.5	417	4	Q96B08
11	624.5	32.7	417	4	Q96E94
12	621	32.5	412	11	P97597
13	531	27.8	279	11	Q9CVD1
14	516.5	27.0	419	6	Q9TV85
15	511	26.7	417	4	Q96GN3
16	511	26.7	417	4	Q96A12

17	489.5	25.6	419	4	Q9B567	Q9B567 homo sapien
18	471.5	24.7	247	4	Q96K29	Q96K29 homo sapien
19	460	24.1	424	5	Q9V187	Q9V187 drosophila
20	459	24.0	1192	5	Q9W475	Q9W475 drosophila
21	451.5	23.6	422	5	Q9V122	Q9V122 drosophila
22	436	22.8	584	5	Q9TZH1	Q9TZH1 caenorhabd
23	418	21.9	440	5	Q9W478	Q9W478 drosophila
24	407.5	21.3	419	5	Q9V121	Q9V121 drosophila
25	402	21.0	424	5	Q9V866	Q9V866 drosophila
26	400.5	21.0	430	5	Q9V186	Q9V186 drosophila
27	399.5	20.9	427	5	Q9U9K2	Q9U9K2 aedes aegypt
28	397	20.9	437	5	Q19121	Q19121 caenorhabd
29	395.5	20.7	540	5	Q9TZH2	Q9TZH2 caenorhabd
30	391	20.5	433	5	Q9V342	Q9V342 drosophila
31	385	20.1	423	5	Q61532	Q61532 drosophila
32	382.5	20.0	581	5	Q23318	Q23318 caenorhabd
33	370	19.4	545	5	Q9N3S6	Q9N3S6 caenorhabd
34	367	19.2	1430	5	Q9VX86	Q9VX86 drosophila
35	366	19.2	457	5	Q9VCM8	Q9VCM8 drosophila
36	362.5	19.0	465	5	Q9VCM6	Q9VCM6 caenorhabd
37	358	18.7	666	5	Q9XU75	Q9XU75 caenorhabd
38	355	18.6	415	5	Q9V564	Q9V564 drosophila
39	349.5	18.3	400	5	Q9V567	Q9V567 drosophila
40	344.5	18.0	354	5	Q9V568	Q9V568 drosophila
41	344.5	18.0	418	5	Q961J8	Q961J8 drosophila
42	337	17.6	312	5	Q9VR23	Q9VR23 drosophila
43	333	17.4	315	4	Q9NR19	Q9NR19 homo sapien
44	324.5	17.0	161	13	Q90W12	Q90W12 brachydanio
45	320	16.7	418	3	Q14418	Q14418 metathizium

ALIGNMENTS

RESULT 1
ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE CARBOXYPEPTIDASE B-LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; AB011869; BAA90475.1; -.
DR HSP; P00730; ZC7C.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; PROPEP_M14.
DR InterPro; IPR000834; Zn_carboxypep.
DR Pfam; PF02244; PROPEP_M14; 1.
DR Pfam; PF00246; Zn_carboxypep; 2.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR PROSITE; PS00765; 40935 MW; BF670B2F7437C1CB CRC64;
SQ SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

NO.

Query Match 98.6%; Score 1884; DB 4; Length 360;
Best Local Similarity 98.3%; Pred. No. 2.2e-162;
Matches 354; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

OY 1 MKLCSLAVLPVIVLFCBOHFAFGSGOVLALPRTSRQOVLOLNTTTEIYVLMQPTAD 60
DB 1 MKLCSLAVLPVIVLFCBOHFAFGSGOVLALPRTSRQOVLOLNTTTEIYVLMQPTAD 60
OY 61 LIYKKQVHFFVNASDVNDVNAKHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
DB 61 LIYKKQVHFFVNASDVNDVNAKHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
OY 121 EOYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLVKSGKQOTAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLVKSGKQOTAKNAIWDGCI 180
OY 121 EOYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLVKSGKQOTAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLVKSGKQOTAKNAIWDGCI 180
OY 181 HAREWISPAFLCWFIGHNRRMRKNSRFFVANNHCIGTDLNSNFVSKHWCCEGASSSCSET 240
DB 181 HAREWISPAFLCWFIGHNRRMRKNSRFFVANNHCIGTDLNSNFVSKHWCCEGASSSCSET 240
OY 241 YCGLYPESEPEVKAASFLLRNINQIKAYISMHSYSHOIVFPYSYRSKSKDHEELSLVA 300
DB 241 YCGLYPESEPEVKAASFLLRNINQIKAYISMHSYSHOIVFPYSYRSKSKDHEELSLVA 300
OY 301 SEAVRAIDKTSKNTRYTHGSESTLYLAPGCGDDWITDGIKYSFNSNPVEKILPLSLK 360
DB 301 SEAVRAIDKTSKNTRYTHGSESTLYLAPGCGDDWITDGIKYSFNSNPVEKILPLSLK 360
RESULT 2
OJ5114 PRELIMINARY: PRT: 423 AA.
AC 015114:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PCPB PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Mallory B.E., Tsai S.P., Henzel W., Dryden D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21835-21838(1991).
DR EMBL; M75106; AAA60042.1; -.
DR HSP; P00730; 1PVT.
DR MEROPS; M14.009; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxpept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF02246; Zn_carboxpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR CHAIN 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ SEQUENCE 423 AA; 48442 MW; 9B383272F6EE79F4 CRC64;

```

Full length

```

DB 121 EOYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLVKSGKQOTAKNAIWDGCI 180
OY 181 HAREWISPAFLCWFIGH-----NRMRK 203
DB 181 HAREWISPAFLCWFIGH-----NRMRK 203
OY 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESPEVKAASFLLRNI 263
DB 241 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESPEVKAASFLLRNI 300
OY 264 NOIKAYISMHSYSHOIVFPYSYRSKSKDHEELSLVA SEAVRAIDKTSKNTRYTHGSE 323
DB 301 NOIKAYISMHSYSHOIVFPYSYRSKSKDHEELSLVA SEAVRAIDKTSKNTRYTHGSE 360
OY 324 TLYLAPGCGDDWITDGIKYSFT 346
DB 361 TLYLAPGCGDDWITDGIKYSFT 383
RESULT 3
OJ5114 PRELIMINARY: PRT: 423 AA.
AC 0961Y4:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B2 (PLASMA).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKELETAL MUSCLE;
RA Strausberg R.;
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC007057; AA07057.1; -.
SQ SEQUENCE 423 AA; 48412 MW; 9B383E03F6EE7CF5 CRC64;

```

```

Query Match 93.9%; Score 1794.5; DB 4; Length 423;
Best Local Similarity 89.3%; Pred. No. 3.6e-154;
Matches 342; Conservative 1; Mismatches 3; Indels 37; Gaps 1;
OY 1 MKLCSLAVLPVIVLFCBOHFAFGSGOVLALPRTSRQOVLOLNTTTEIYVLMQPTAD 60
DB 1 MKLCSLAVLPVIVLFCBOHFAFGSGOVLALPRTSRQOVLOLNTTTEIYVLMQPTAD 60
OY 61 LIYKKQVHFFVNASDVNDVNAKHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
DB 61 LIYKKQVHFFVNASDVNDVNAKHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
OY 121 EOYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLVKSGKQOTAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLVKSGKQOTAKNAIWDGCI 180
OY 121 EOYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLVKSGKQOTAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLVKSGKQOTAKNAIWDGCI 180
OY 181 HAREWISPAFLCWFIGH-----NRMRK 203
DB 181 HAREWISPAFLCWFIGH-----NRMRK 203
OY 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESPEVKAASFLLRNI 263
DB 241 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESPEVKAASFLLRNI 300
OY 264 NOIKAYISMHSYSHOIVFPYSYRSKSKDHEELSLVA SEAVRAIDKTSKNTRYTHGSE 323
DB 301 NOIKAYISMHSYSHOIVFPYSYRSKSKDHEELSLVA SEAVRAIDKTSKNTRYTHGSE 360
OY 324 TLYLAPGCGDDWITDGIKYSFT 346
DB 361 TLYLAPGCGDDWITDGIKYSFT 383

```

```

RESULT 4
ID 09JHH6 PRELIMINARY; PRT; 422 AA.
AC 09JHH6;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE R (THROMBIN-ACTIVATABLE FIBRINOLYSIS INHIBITOR)
DE (111003P04RIK PROTEIN).
GN CPB2 OR TAFI OR 1110032P04RIK.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20341711; PubMed=10878383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Odate K., Okada N.,
RA Campbell W., Okada H.;
RT "pro-carboxypeptidase R is an acute phase protein in the mouse,
RT whereas carboxypeptidase N is not.";
RL J. Immunol. 165:1053-1058(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20201996; PubMed=10739389;
RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
RT inhibitor.";
RL Thromb. Haemost. 83:297-303(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsch G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilmink L.,
RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AB021968; BAB03402.1; -
DR EMBL; AF164524; AAF62385.1; -
DR EMBL; AK004045; BAB23141.1; -
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -.
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

```

Query Match 80.0%; Score 1528; DB 11; Length 422;
 Best Local Similarity 76.2%; Pred. No. 5e-130;
 Matches 292; Conservative 24; Mismatches 29; Indels 38; Gaps 2;

QY 1 MKLCSLAVLPYIVLFCQHAFQSGOVLALPRTSRQVQLQNLTTTYEVLVMPQVTAD 60

```

DB 1 MKLHGLILVAIILY-EGHGFQSGOVLASLPRTSRQVQLQNLTTTYEVLVMPQVTAE 59
QY 61 LIYKKQVHFVNASDVNDVNVKAHLNVSGIPCSYVLADVEDILQOQISNDVSPRASASY 120
DB 60 FIEKKKEHFVNASDVSDVSKAHLNVSRIPNVLMNVEDILQOQTRNDVSPRASASY 119
QY 121 EGYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKYPYLVLKSGKEQTANAIWIDCGI 180
DB 120 EGYHSLNEIYSWIEFIERHPDMLQKIYIGSSFERYPLVLKSGKORIKNAIWDGCI 179
QY 181 HAREWISPAFLMFIH-----NRMKR 203
DB 180 HAREWISPAFLMFIHGYVTOFHGREKENTYRLRHVDFYIMPVNWDYDYTWKKNRMWRK 239
QY 204 NRSFYANNHCIGDNLNPFNSKHWCCEGASSSCSEYTCGIPSEPEVKVASFRLRNI 263
DB 240 NRSANNNRCVGTDLNRFNSKHWCCEGASSSCSEYTCGIPSEPEVKVADFLRNI 299
QY 264 NQIKAYISMYSQHIYFPYSYTRSKSDHEELSLVSEAVRAIDKTSKNTRYTHGSGE 323
DB 300 DHIKAYISMYSQQLIFPYSYTRSKSDHEELSLVSEAVRAIESINKNTRYTHGSGE 359
QY 324 TLYLAPGGDDWYIDLGIKYSFT 346
DB 360 SLYLAPGGSDWYIDLGIKYSFT 382

```

```

RESULT 5
ID 09QZFO PRELIMINARY; PRT; 422 AA.
AC 09QZFO;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE U.
GN CPB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAF1J;
RX EMBL; AF186188; AAF00528.1; -
RA He Y.C., Broze G.;
RT "Isolation and characterization of mouse liver carboxypeptidase B
RT gene.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RA EMBL; AF186188; AAF00528.1; -
RA HSSP; P00730; 5CPA.
RA MEROPS; M14.009; -.
RA MGD; MGI:1891837; Cpb2.
RA InterPro; IPR003146; Propep_M14.
RA InterPro; IPR000834; Zn_carboxypept.
RA Pfam; PF02244; Propep_M14; 1.
RA Pfam; PF00246; Zn_carboxypept; 1.
RA PRINTS; PR00765; CRBOXYPTASEA.
SQ SEQUENCE 422 AA; 48832 MW; B53FFB09943954E CRC64;

```

Query Match 79.6%; Score 1522; DB 11; Length 422;
 Best Local Similarity 76.0%; Pred. No. 1.7e-129;
 Matches 291; Conservative 24; Mismatches 30; Indels 38; Gaps 2;

QY 1 MKLCSLAVLPYIVLFCQHAFQSGOVLALPRTSRQVQLQNLTTTYEVLVMPQVTAD 60

DB 1 MKLHGLILVAIILY-EGHGFQSGOVLASLPRTSRQVQLQNLTTTYEVLVMPQVTAE 59

QY 61 LIYKKQVHFVNASDVNDVNVKAHLNVSGIPCSYVLADVEDILQOQISNDVSPRASASY 120

DB 60 FIEKKKEHFVNASDVSDVSKAHLNVSRIPNVLMNVEDILQOQTRNDVSPRASASY 119

QY 121 EGYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKYPYLVLKSGKEQTANAIWIDCGI 180

DB 120 EGYHSLNEIYSWIEFIERHPDMLQKIYIGSSFERYPLVLKSGKORIKNAIWDGCI 179

```
Db 120 EQHSLNEIYSWIEFTEFRHPDMLTKIHGSSFEKPYLYLVKSGKQRIKNAIWDGCI 179
QY 181 HAREWISPAFCLEWFIH-----NRMRK 203
Db 180 HAREWISPAFCLEWFIH-----NRMRK 203
QY 204 NRSFYANNHCIGDNLNENFSKHKCEGASSSCSEFYGLYPESEPEKAVASFLRNT 263
Db 240 NRSYAHNNRCVGTDLNENFASKHCEGASSSCSEFYGLYPESEPEKAVADFLRNT 299
QY 264 NQIKAYISMSYSQHIYFPYSTRSKSDHELSLVAEVAIRADKTSKRTYTHGSGE 323
Db 300 DHIKAYISMSYSQHIYFPYSTRSKSDHELSLVAEVAIRADKTSKRTYTHGSGE 359
QY 324 TYLAPGSGDDMIYDLGIKYSFT 346
Db 360 SYLAPGSGDDMIYDLGIKYSFT 382
```

RESULT 6

```
Q9EOV9 PRELIMINARY; PRT; 422 AA.
AC Q9EOV9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PRE-PROCARBOXYPEPTIDASE R.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=20471387; PubMed=11021404;
RA Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
Okada N., Okada H.;
RT "Molecular cloning and partial characterization of rat
RT procarboxypeptidase R and carboxypeptidase N.";
RL Microbiol. Immunol. 44:719-728(2000).
DR EMBL: AB042598; BAB18617.1; -.
DR HSRP: P00730; ZC7C.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxept; 1.
DR PRINTS: PR00765; CRBOXYPTRASEA.
KW Carboxypeptidase.
SQ SEQUENCE 422 AA; 48826 MW; FFFD32A51A9366C8 CRC64;
```

Query Match 78.4%; Score 1499; DB 11; Length 422;
Best Local Similarity 74.7%; Pred. No. 2.1e-127;
Matches 286; Conservative 26; Mismatches 33; Indels 38; Gaps 2;

```
QY 1 MKLCSLAVLVPIVFCGQHVAFQSGOVLALPRTSRQVQVQLNTTYYEIVLMQPYTAD 60
Db 1 MKLYGLVLAIVLIFCQHVAFQSGOVLALPRTSRQVQVQLNTTYYEIVLMQPYTAE 59
QY 61 LIVKKQVHFVNASDVNDYKAHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
Db 60 FIEKKKEVHFVNASDVNDYKAHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 119
QY 121 EQHSLNEIYSWIEFTEFRHPDMLTKIHGSSFEKPYLYLVKSGKQRIKNAIWDGCI 180
Db 120 EQHSLNEIYSWIEFTEFRHPDMLTKIHGSSFEKPYLYLVKSGKQRIKNAIWDGCI 179
QY 181 HAREWISPAFCLEWFIH-----NRMRK 203
Db 180 HAREWISPAFCLEWFIH-----NRMRK 203
QY 204 NRSFYANNHCIGDNLNENFSKHKCEGASSSCSEFYGLYPESEPEKAVASFLRNT 263
Db 240 NRSYAHNNRCVGTDLNENFASKHCEGASSSCSEFYGLYPESEPEKAVADFLRNT 299
```

```
Db 240 NRSYAHNNRCVGTDLNENFASKHCEGASSSCSEFYGLYPESEPEKAVADFLRNT 299
QY 264 NQIKAYISMSYSQHIYFPYSTRSKSDHELSLVAEVAIRADKTSKRTYTHGSGE 323
Db 300 DHIKAYISMSYSQHIYFPYSTRSKSDHELSLVAEVAIRADKTSKRTYTHGSGE 359
QY 324 TYLAPGSGDDMIYDLGIKYSFT 346
Db 360 SYLAPGSGDDMIYDLGIKYSFT 382
```

RESULT 7

```
Q9NTI8 PRELIMINARY; PRT; 198 AA.
AC Q9NTI8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BA139H14.2 (CARBOXYPEPTIDASE B2 (PLASMA)) (FRAGMENT).
GN CPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RA Tromans A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL137141; CAB92622.1; -.
DR HSRP: P00730; ZC7C.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxept; 1.
DR PRINTS: PR00765; CRBOXYPTRASEA.
KW Carboxypeptidase.
FT NOK_TER 198
SQ SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;
```

Query Match 53.8%; Score 1029; DB 4; Length 198;
Best Local Similarity 99.5%; Pred. No. 2.7e-85;
Matches 196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 MKLCSLAVLVPIVFCGQHVAFQSGOVLALPRTSRQVQVQLNTTYYEIVLMQPYTAD 60
Db 1 MKLYGLVLAIVLIFCQHVAFQSGOVLALPRTSRQVQVQLNTTYYEIVLMQPYTAE 59
QY 61 LIVKKQVHFVNASDVNDYKAHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
Db 61 LIVKKQVHFVNASDVNDYKAHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
QY 121 EQHSLNEIYSWIEFTEFRHPDMLTKIHGSSFEKPYLYLVKSGKQRIKNAIWDGCI 180
Db 121 EQHSLNEIYSWIEFTEFRHPDMLTKIHGSSFEKPYLYLVKSGKQRIKNAIWDGCI 180
QY 181 HAREWISPAFCLEWFIH-----NRMRK 203
Db 181 HAREWISPAFCLEWFIH-----NRMRK 203
QY 204 NRSFYANNHCIGDNLNENFSKHKCEGASSSCSEFYGLYPESEPEKAVASFLRNT 263
Db 240 NRSYAHNNRCVGTDLNENFASKHCEGASSSCSEFYGLYPESEPEKAVADFLRNT 299
```

OX NCBI_TaxID=8724;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
 RT "Screening of Bothrops jararaca pancreas cDNA library";
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF190274; AAF01344.1; -
 DR HSP: P09955; INSA.
 DR MEROPS: M14.003; -
 DR InterPro: IPR003146; Propep.M14.
 DR InterPro: IPR000834; Zn_carboxypept.
 DR Pfam: PF02244; Propep.M14; 1.
 DR Pfam: PF00246; Zn_carboxypept. 1.
 DR PRINTS: PR00765; CARBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 SQ SEQUENCE 416 AA; 47723 MW; 4F999854DD7287A7 CRC64;
 Query Match 36.7%; Score 700.5; DB 13; Length 416;
 Best Local Similarity 38.5%; Pred. No. 4,5e-55;
 Matches 142; Conservative 65; Mismatches 119; Indels 43; Gaps 5;
 QY 15 FCGHAFAGSGOVLALPRTSRQOVLQNLTTTTEIVLMQPTADLIYKKQHFYNA 74
 Db 14 FAFTYHREDGEKYVRYTPNEDEVFLNTLANIVQDFMRPSVELYKAKMTDFRTEA 73
 QY 75 SDVNTKALHNVSGIPCSVLLADVEDLIQOQISNDTVSPRASAYEYQYSLNEIYSWIE 134
 Db 74 DRCSEVESILQSGNLNELLIDNLQAVLDRLDNLH---RTAGVNEKYMSWEKIDMTA 130
 QY 135 FTERHPDMLTKHISSEFKYLYLVKSGKQOTAKNAIWDGHHARWISPAFLMF 194
 Db 131 DIANENPSLVSRLQIGTFGRPMPLKY-GKPGVNKKAIFIDCGFARWISPAFLMF 189
 QY 195 I-----GHNRMKRKRSFYANNHCIGTD 217
 Db 190 VRAVAVTYKETITMOLINKLDFYLPVLNIDGYVSMKOSRMKRKTRVNASGTCIGTD 249
 QY 218 LNSNFVSKHNCBEGASSSCSETTCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYSQ 277
 Db 250 PNNNF-DAAMCVSGASNPSETTCGSKPESEKETALADFIIRNRRIIDAYLIHSYSQ 308
 QY 278 HLYFPSTYRSKSDHELSVASEAVRAIDKSKTRITHGSEFLYLAPEGGDWIT 337
 Db 309 MLYPSTYTYDLSNNKLNISAKIAIRL-KVLFTEYTGPGCAATYPAAGSDDMAY 367
 QY 338 DLGIXYSFT 346
 Db 368 DQGIKYAFT 376
 RESULT 9
 ID Q9XSP3 PRELIMINARY; PRT; 416 AA.
 AC Q9XSP3;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PROCARBOXYPEPTIDASE B PRECURSOR (EC 3.4.17.2).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PANCREAS;
 RX MEDLINE=99321926; PubMed=10391940;
 RA Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
 RA Hersberger C., Ayiles F.;
 RT "Mapping the Pro-region of carboxypeptidase B by protein engineering.
 RT Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
 RL J. Biol. Chem. 274:19925-19933(1999).

DR EMBL: AJ133775; CAB46991.1; -
 DR HSP: P09955; IPBA.
 DR MEROPS: M14.003; -
 DR InterPro: IPR003146; Propep.M14.
 DR InterPro: IPR000834; Zn_carboxypept.
 DR Pfam: PF02244; Propep.M14; 1.
 DR Pfam: PF00246; Zn_carboxypept. 1.
 DR PRINTS: PR00765; CARBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 DR SIGNAL: Carboxypeptidase; Hydroxylase.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 416 PROCARBOXYPEPTIDASE B.
 SQ SEQUENCE 416 AA; 47381 MW; 5FAFCEED7B45ABEA CRC64;
 Query Match 33.8%; Score 646; DB 6; Length 416;
 Best Local Similarity 36.6%; Pred. No. 3.9e-50;
 Matches 140; Conservative 64; Mismatches 134; Indels 44; Gaps 7;
 QY 3 LCSIAVLVPIVLFCEQHVFA-FQSGOVLALPRTSRQOVLQNLTTTTEIVLMQPTADL 61
 Db 1 MIAFLILVTVTLASAHSGEFGKEKVRVNVDENDISLHSLASTRLQDFMRKPDVYQ 60
 QY 62 IYKKQYHFYNASVDVNVKALHNVSGIPCSVLLADVEDLIQOQISNDTVSPRASAYE 121
 Db 61 IKPSTVDFRKADILADVEDLEQNELOVEVLINNLRSVLAOFDSRV--RTGHSYE 117
 QY 122 QYHSLNEIYSWIEFTRHPDMLTKHISSEFKYLYLVKSGKQOTAKNAIWDGHH 181
 Db 118 KYNWETIEATKQVTSNPDLISRTAIGTGLNNIYLLKY-GKPGVNKKAIFIDCGF 176
 QY 182 AREWISPAFLMF-----GH-----NRMRKN 204
 Db 177 AREWISPAFLMF-----GH-----NRMRKN 204
 QY 205 RSFYANNHCIGTDLNSNFVSKHNCBEGASSSCSETTCGLYPESEPEVKAVASFLRR 264
 Db 237 RSTNAGTTCIGTDNPNRF-DAGMCTGASTDPCDETCGSAASEKETKALADFIIRNLS 295
 QY 265 QIKAYISMHSYSHIIVPYSTYRSKSDHELSVASEAVRAIDKSKTRITHGSEFLY 324
 Db 296 STAYLIHSTQMLTPYSTDYKLPENNALNLAAKVEL-ATLYGTITTYGPGATT 354
 QY 325 LYLAPEGGDWITYDLGIXYSFT 346
 Db 355 IYPAAGSDDMAYDQGIKYSET 376
 RESULT 10
 ID Q96B08 PRELIMINARY; PRT; 417 AA.
 AC Q96B08;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE HYPOHETICAL 47.4 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ADRENAL CARCINOMA;
 RA Strausberg R.;
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC015338; AAH15338.1; -
 DR Hypothetical protein.
 SQ SEQUENCE 417 AA; 47367 MW; EBBB98B27F5D5AF9 CRC64;
 Query Match 33.5%; Score 641; DB 4; Length 417;
 Best Local Similarity 35.3%; Pred. No. 1.1e-49;

```

QY      205 KSFVANNHCIGTDLNSNFVKSHWCEGASSSCSEFTYGLYPESEPEKAVASFLRRIN
       11 :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      237 RSKNNSKCIGTDLNRNF--NASMNSIPMTNPDCADNYRGSAPESKEKKAATNIRSHLN
       11 :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY      265 QAKVAISMYSQOHIVFPFYSTSRKSCKCHELSLVASAANAIDIKTSKNNTYTTHGHSET
       11 :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      296 ELKVITTFHSYSQMILFFPGYTSKLPPHHEDIAKVIAGTDVL-STREYRYIYGPIEST
       11 :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY      325 LYLAFGGGDWTLYDGIKRSF 345
       11 :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      355 IYPISGSSLDMAVDLGIKHTE 375
       11 :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 12
P97597          PRELIMINARY;             PRT;           412 AA.
P97597;
AC              P97597;
DT              01-MAY-1997 (TREMBLrel_03, Created)
DT              01-MAY-1997 (TREMBLrel_03, Last sequence update)
DI              01-DEC-2001 (TREMBLrel_19, last annotation update)
DE              MAST CELL CARBOXYPEPTIDASE A PRECURSOR (FRAGMENT).
GN              R-CPA.
OS              Rattus norvegicus (Rat.).
OC              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN              NCBI_maxid-10116;
RP              [1]
SQ              SEQUENCE FROM N.A.
RC              STRAIN=SPRACHE DAMLEY;
RX              MEDLINE=97149430; PubMed=8996238;
RA              Lutzeischwab C., Pejler G., Aveskogh M., Hellman L.;
RT              "Secretory granule proteases in rat mast cells. Cloning of 10
               different serine proteases and a carboxypeptidase A from various rat
               mast cell populations." ;
RL              J. Exp. Med. 185:13-29(1997).
DR              EMBL: U67914; FAMB48267.1; -.
DR              HSSP: P09955; INSA.
DR              MEROPS: M14.010; -.
DR              InterPro: IPR003146; Propep_M14.
DR              InterPro: IPR000834; zn_carbopept.
DR              Pfam: PF02244; Propep_M14; 1.
DR              Pfam: PF00246; zn_carpoepet; 1.
DR              PRINTS: PR00765; CRBOXYPTASER.
DR              PROSITE: PS00133; CARBOXYPEPT_ZN_1; 1.
DR              PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW              Signal; Carboxypeptidase.
FT              SIGNAL                     1
FT              NON_TER                    1
FT              CHAIN                      1
FT              SIGNAL                     1
FT              CHAIN                      1
SQ              SEQUENCE 412 AA; 47944 MW; DAB59555FC49137D CRC64;

```

	Query Match	32.5%	Score 621;	DB 11;	Length 412;
	Best Local Similarity	37.28%;	Pred. No. 7e-48;		
	Matches 134;	Conservative 55;	Mismatches 129;	Indels 42;	Gaps 6;
OY	23	FQSGGVALAALPRSRVGYLQMLTITTEIYLVMOPRTADLYKKKQVHFYNASDVNKA	82		
		: : :	:	:	:
Db	16	FDKEKFRAKLQDEKASTILKNLTOTIELDWPYPDAIHDAVNMTVDFFERTENESOTIS	75		
OY	83	HLNVSGIPCSVLADVEDLILOOQISINDPVSPPASASYEQGHSLNEIYSWIEFTIRRHQ	142		
		: : :	:	:	:
Db	76	TLEQHMDMEIILLINLOEIDRQF--DYKEELAGRHSHAYANOMNRKYSTMLEGVAKHE	133		
OY	143	MLTKHHIGGSFEKPYLVLYLKVGSKQTAKNAIWDICSIHAREMISPAFLCMFT-----	195		
		: : :	:	:	:
Db	134	MSRKITGSTVEDNPYLAKTI-GRKGGERKAIFMDCGIHAREVSPAFQCFMYQAOKSY	192		
OY	196	GHN-----BMRKRNSFYANNNICITDLSNFVSFK	225		
		: : :	:	:	:
Db	193	GKKRIWTKLDDRMNFYYLPVFENVGDYIMSWTKDDMRMKRNSKNPNSTCIGTDIDLRNF-DY	251		

```

Db 252 SWDSSPNTDPCLSYVRGPAPESEKETKAVNTIRSHINIKAYITPHSHSOMLIPYGY 311
QY 286 TRSKSDHEELSLVASEAVRAIDKTSKNTYTHGSESLYLAPGGDDMIYDLGIKYSF 345
Db 312 TIKLPNHODLKAVARIARDVL-SSRYEPRYIGYPIASTIKYKSGSLDMAYDGLKHTF 370

RESULT 13
Q9CYD1 PRELIMINARY; PRT; 279 AA.
AC 09CYD1
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 221008M23RIK PROTEIN (FRAGMENT).
GN 221008M23RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kledaschmann W., Gaasterland T., Gissi C., King B., Kochiba H.,
RA Kuenl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quakebush J.,
RA Schiml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
RA Lyons F., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whittaker C., Wilmink L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK008678; BAB25826.1;
DR HSSP: P09955; INSA.
DR MGD: MGI:1923953; 221008M23RIK.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CARBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; UNKNOWN_1.
FT NON_TER 279
FT SEQUENCE 279 AA; 31970 MW; 29468AFA0B67E0AA CRC64;

Query Match 27.8%; Score 531; DB 11; Length 279;
Best Local Similarity 45.6%; Pred. No. 5, 7e-40;
Matches 103; Conservative 31; Mismatches 52; Indels 40; Gaps 4;
QY 158 LVYKVSQGEQRAKNAIMIDCGIHAREWISPAFCIMFI----- 195
Db 1 MYVLKI-GKDRPKPAFFIDCGFHAREWISPAFCOMFVEAVTYKQEIHKRLDELDF 59
QY 196 -----GHNRMKRNRSFYANNHCIGTDLNSNEVSKHMGEGASSSCSET 240
Db 60 YVLPVNVNDICYYTAKDKMMKRTKRTSTTAGSSCGFVDPNRNF-DAQCEGASRSFCSDT 118
QY 241 YGGLPPESEPEYKAAVASFRLRNINQIKAYISMHSQHIYFPYSTRSKSDHEELSLVA 300
Db 119 YGPFPESEKETKALADFIKONLSIKAYILVHSYQMLPYSDYKLPENEEELNALV 178
QY 301 SFAVRAIDKTSKNTYTHGSESLYLAPGGDDMIYDLGIKYSF 346

```

```

Db 179 KGAKEEL-STLHGKTYTYGPGATTYPAAGSGSDMAYDGLKYSF 223

RESULT 14
Q9TV85 PRELIMINARY; PRT; 419 AA.
AC 09TV85
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE A1 (EC 3.4.17.1).
GN PCPAl.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9823;
RN 11
RP SEQUENCE FROM N.A.
RC MEDLINE=99192816; PubMed=10092856;
RX Darnis S., Juge N., Marino C., Aviles F.X., Fulgiver A., Chaux J.C.,
RA Guo X.J.;
RT "Cloning, sequencing and functional expression of a cDNA encoding
RT porcine pancreatic preprocarboxypeptidase A1.";
RL Eur. J. Biochem. 259:719-725(1999).
DR EMBL: AF076222; AAD17690.1;
DR HSSP: P09954; 1PCA.
DR MEROPS: M14.001;
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CARBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase.
SQ SEQUENCE 419 AA; 47235 MW; 84B4CB57B714FC1 CRC64;

Query Match 27.0%; Score 516.5; DB 6; Length 419;
Best Local Similarity 33.8%; Pred. No. 2, 1e-38;
Matches 132; Conservative 63; Mismatches 133; Indels 63; Gaps 12;
QY 3 LCSIATVLPVLFCEQHVAFQSGVLAALPRSRQVYVLTNTTYEIVL-WQ-PVTA 59
Db 4 LLIFFSVLLGGVLAKED---FVGHQVLRISVDDEAQYQVKLEDELHLDLDFWRGPAR 59
QY 60 DLYKKQVAFPFVNASVDVNAKHLNAGIPCSVLLADVEDLI---QQQISNDTVSPRAS 116
Db 60 GFPIDVR---VFPSIQVKKVFLFAHGIRYTIMEDVOLLDEDEQOMFASQGRART 114
QY 117 ASY-YEQYHSLNEYSWIEITERHERPMLTKRIHGSFEFYPLVYLKVSQGEQRAK--- 172
Db 115 STRYATYHLEELIYDPMDLVAEHPLQVSKLQIGSSIEGRPIYLVKFS---TGNNRP 170
QY 173 AIWIDCGIHAREWISPAFCIMFI----- 195
Db 171 AIWIDTGHISREWYTAQSGWPAFKITREDYGQDPATAIIDNDITFLEIYNPDGFAFTH 230
QY 196 GHNRMKRNRSFYANNHCIGTDLNSNFVSKHMGEGASSSCSETTYGGLPSEPEYKAY 255
Db 231 SENMNMKRTSKRTSGSCVCPVDPNRNF-DAGFGAGASSNPGSETYHGKFPNSEVEVKSI 289
QY 256 ASFLRRNINQIKAYISMHSQHIYFPYSTRSKSDHEELSLVAFAVRAIDKTSKNT 315
Db 290 VDEVNDHGN-ITAFISHISQILLIYPYKTEAPADKDELQISKAVAL-ITSLYGTK 347
QY 316 YTHGSESLYLAPGGDDMIYDLGIKYSF 346
Db 348 FQYGSITTTIYQASGRTIDWTYNGIKYSFS 378

RESULT 15

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:39:53 ; Search time 28.95 Seconds
(without alignments)
303.738 Million cell updates/sec

Title: US-09-980-881-2

Perfect score: 1911

Sequence: 1 MKLCSLAVLVPIVIFCEQHV.....IKYFTSNPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 segs, 24425594 residues

231628

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgcn2_6/ptodata/1/1aa/5A_COMB.pep:*

2: /cgcn2_6/ptodata/1/1aa/5B_COMB.pep:*

3: /cgcn2_6/ptodata/1/1aa/5A_COMB.pep:*

4: /cgcn2_6/ptodata/1/1aa/5B_COMB.pep:*

5: /cgcn2_6/ptodata/1/1aa/PTUS_COMB.pep:*

6: /cgcn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1799.5	94.2	423	1	US-07-649-591B-3
2	1799.5	94.2	423	1	US-08-277-540-3
3	1799.5	94.2	423	1	US-08-430-787A-3
4	1799.5	94.2	423	2	US-08-869-057-2
5	640.5	33.5	404	1	US-08-696-139-2
6	639	33.4	415	2	US-08-860-882A-57
7	628.5	32.9	417	1	US-07-649-591B-7
8	628.5	32.9	417	1	US-08-277-540-7
9	628.5	32.9	417	1	US-08-430-787A-7
10	623	32.6	716	4	US-09-171-945-125
11	619.5	32.4	417	1	US-07-649-591B-6
12	619.5	32.4	417	1	US-08-277-540-6
13	619.5	32.4	417	1	US-08-430-787A-6
14	604	31.6	307	2	US-08-782-760-6
15	604	31.6	307	5	PCT-US96-00995-6
16	603.5	31.6	396	1	US-07-649-591B-4
17	603.5	31.6	396	1	US-08-277-540-4
18	603.5	31.6	396	1	US-08-430-787A-4
19	576	30.1	306	1	US-08-696-139-4
20	553	28.9	417	1	US-07-649-591B-8
21	553	28.9	417	1	US-08-277-540-8
22	553	28.9	417	1	US-08-430-787A-8
23	543	28.4	613	4	US-09-171-945-113
24	528.5	27.7	419	1	US-07-649-591B-5
25	528.5	27.7	419	1	US-08-277-540-5
26	528.5	27.7	419	1	US-08-430-787A-5
27	513	26.8	417	4	US-08-640-906-4

28	513	26.8	417	4	US-09-395-936-4	Sequence 4, Appl
29	512	26.8	417	4	US-08-640-906-18	Sequence 18, Appl
30	512	26.8	417	4	US-09-395-936-18	Sequence 18, Appl
31	487.5	25.5	419	4	US-08-640-906-2	Sequence 2, Appl
32	487.5	25.5	419	4	US-09-395-936-2	Sequence 2, Appl
33	482.5	25.2	419	4	US-08-640-906-17	Sequence 17, Appl
34	482.5	25.2	419	4	US-09-395-936-17	Sequence 17, Appl
35	167	8.7	37	1	US-07-649-591B-1	Sequence 1, Appl
36	167	8.7	37	1	US-08-277-540-1	Sequence 1, Appl
37	167	8.7	37	1	US-08-430-787A-1	Sequence 1, Appl
38	118.5	6.2	706	1	US-08-484-105-16	Sequence 16, Appl
39	118.5	6.2	706	1	US-08-484-106-16	Sequence 16, Appl
40	100.5	5.3	95	2	US-08-782-760-3	Sequence 3, Appl
41	100.5	5.3	95	5	PCT-US96-00995-3	Sequence 3, Appl
42	95	5.0	734	4	US-08-706-216-2	Sequence 2, Appl
43	94.5	4.9	1683	3	US-08-755-587-183	Sequence 183, App
44	90.5	4.7	346	4	US-09-073-297-2	Sequence 2, Appl
45	90.5	4.7	1726	2	US-08-609-049A-30	Sequence 30, Appl

ALIGNMENTS

```

RESULT 1
US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: NO. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: pathn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; US-07-649-591B-3

Query Match 94.2%; Score 1799.5; DB 1; Length 423;
Best Local Similarity 89.6%; Pred. No. 4.2e-187;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;
Oy 1 MKLCSLAVLVPIVIFCEQHVAFQSGOVLALPTSRQOVVLQNTTTEIVLMQPTAD 60
Db 1 MKLCSLAVLVPIVIFCEQHVAFQSGOVLALPTSRQOVVLQNTTTEIVLMQPTAD 60

```

```

QY 61 LIYKKQVHFVNASDVNDVKAHLNVSIGIPCSVLLADVEDL100QISNDTVSPRASASY 120
DB 61 LIYKKQVHFVNASDVNDVKAHLNVSIGIPCSVLLADVEDL100QISNDTVSPRASASY 120
QY 121 EOYHSLNEIYSWIEFTEHRPDLTKIHIGSSFEKYPYLVLKVSKEQTAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFTEHRPDLTKIHIGSSFEKYPYLVLKVSKEQTAKNAIWDGCI 180
QY 181 HAREWISPAFCIMFIGH-----NRMMRK 203
DB 181 HAREWISPAFCIMFIGHITQFYIGQYTNLLRLVDFYVMPVNVVDGYDSMKKRRMMRK 240
QY 204 NRSFYANNHCIGTDLNSNFKHMCCEGASSSCSEYTCGLYPESEPEYKAVASFLRRNI 263
DB 241 NRSFYANNHCIGTDLNSNFKHMCCEGASSSCSEYTCGLYPESEPEYKAVASFLRRNI 300
QY 264 NOIKAYISMHSYQHIYVFPYSYTRSKSKDHEELSLVASEAVRAIDTSKNTRYTHGHGSE 323
DB 301 NOIKAYISMHSYQHIYVFPYSYTRSKSKDHEELSLVASEAVRAIDTSKNTRYTHGHGSE 360
QY 324 TLYLAPGGDDMIYDLGIRKYSFT 346
DB 361 TLYLAPGGDDMIYDLGIRKYSFT 383

```

RESULT 2

```

US-08-277-540-3
Sequence 3, Application US/08277540
Patent No. 5474901

```

GENERAL INFORMATION:

```

APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277, 540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435

```

PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91

```

ATTORNEY/AGENT INFORMATION:

```

NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881

```

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 3:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-277-540-3

```

```

Query Match          94.2% Score 1799.5; DB 1; Length 423;
Best local Similarity 89.6% Pred No 4.2e-187;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

```

```

QY 1 MKLCSLAVLPVILFCEQHVAFQSGQVLAALPRTSRQOVQLONTTIEIYLMQPVYAD 60
DB 1 MKLCSLAVLPVILFCEQHVAFQSGQVLAALPRTSRQOVQLONTTIEIYLMQPVYAD 60
QY 61 LIYKKQVHFVNASDVNDVKAHLNVSIGIPCSVLLADVEDL100QISNDTVSPRASASY 120
DB 61 LIYKKQVHFVNASDVNDVKAHLNVSIGIPCSVLLADVEDL100QISNDTVSPRASASY 120
QY 121 EOYHSLNEIYSWIEFTEHRPDLTKIHIGSSFEKYPYLVLKVSKEQTAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFTEHRPDLTKIHIGSSFEKYPYLVLKVSKEQTAKNAIWDGCI 180
QY 181 HAREWISPAFCIMFIGH-----NRMMRK 203
DB 181 HAREWISPAFCIMFIGHITQFYIGQYTNLLRLVDFYVMPVNVVDGYDSMKKRRMMRK 240
QY 204 NRSFYANNHCIGTDLNSNFKHMCCEGASSSCSEYTCGLYPESEPEYKAVASFLRRNI 263
DB 241 NRSFYANNHCIGTDLNSNFKHMCCEGASSSCSEYTCGLYPESEPEYKAVASFLRRNI 300
QY 264 NOIKAYISMHSYQHIYVFPYSYTRSKSKDHEELSLVASEAVRAIDTSKNTRYTHGHGSE 323
DB 301 NOIKAYISMHSYQHIYVFPYSYTRSKSKDHEELSLVASEAVRAIDTSKNTRYTHGHGSE 360
QY 324 TLYLAPGGDDMIYDLGIRKYSFT 346
DB 361 TLYLAPGGDDMIYDLGIRKYSFT 383

```

RESULT 3

```

US-08-430-787A-3
Sequence 3, Application US/08430787A
Patent No. 5593674

```

GENERAL INFORMATION:

```

APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430, 787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514

```

PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 08/277, 540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91

```

ATTORNEY/AGENT INFORMATION:

```

NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616

```

REFERENCE/DOCKET NUMBER: 689D1C1D1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-1896
 TELEFAX: 415/952-9881
 TELEX: 910/371-7168
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 423 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-430-787A-3

Query Match 94.2%; Score 1799.5; DB 1; Length 423;
 Best Local Similarity 89.6%; Pred. No. 4.2e-187;
 Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

QY 1 MKLCSLAVLPVILFCEQHFVAFQSGOVLALPRTSRQVQLNLTYYEIVLMQPTAD 60
 DB 1 MKLCSLAVLPVILFCEQHFVAFQSGOVLALPRTSRQVQLNLTYYEIVLMQPTAD 60
 QY 61 LIYKKQVHFVNASDVNDKKAHLNVSIGIPCSVLADVEDLIQOQISNDTVSPRASASY 120
 DB 61 LIYKKQVHFVNASDVNDKKAHLNVSIGIPCSVLADVEDLIQOQISNDTVSPRASASY 120
 QY 121 EGYHSLNEIYSWIEFTEHPDMLTKIHIGSFEEKYPLYLVKSGKEQTAKNAIWDGCI 180
 DB 121 EGYHSLNEIYSWIEFTEHPDMLTKIHIGSFEEKYPLYLVKSGKEQTAKNAIWDGCI 180
 QY 181 HAREWISPAFLMFIGH-----NRWRK 203
 DB 181 HAREWISPAFLMFIGHITQFYGIIGYTNLRLVDFYVMPVNVVDGYDSMKKNRMWRK 240
 QY 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSEYTCGLYPESEPEVKAASFRLRNI 263
 DB 241 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSEYTCGLYPESEPEVKAASFRLRNI 300
 QY 264 NQIKAYISMHSYQHIVFPYSTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGSGE 323
 DB 301 NQIKAYISMHSYQHIVFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSGE 360
 QY 324 TLYLAPGGDDWIVDGIKYSFT 346
 DB 361 TLYLAPGGDDWIVDGIKYSFT 383

RESULT 4
 US-08-869-057-2
 Sequence 2, Application US/08869057
 Patent No. 5983562
 GENERAL INFORMATION:
 APPLICANT: Morser, Michael J
 APPLICANT: Nagashima, Mariko
 TITLE OF INVENTION: Method of Detecting Thrombotic Disease
 TITLE OF INVENTION: Risk
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Berlex Biosciences Legal Department
 STREET: 15049 San Pablo Avenue
 CITY: Richmond
 STATE: California
 COUNTRY: USA
 ZIP: 94804-0099
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/869,057
 FILING DATE: 03-JUN-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:

NAME: Washlien, Wendy L
 REGISTRATION NUMBER: 36,301
 REFERENCE/DOCKET NUMBER: 51509AUSM1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 510-262-5411
 TELEFAX: 510-262-7095
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 423 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 TISSUE TYPE: Plasma
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 23..401
 US-08-869-057-2

Query Match 94.2%; Score 1799.5; DB 2; Length 423;
 Best Local Similarity 89.6%; Pred. No. 4.2e-187;
 Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

QY 1 MKLCSLAVLPVILFCEQHFVAFQSGOVLALPRTSRQVQLNLTYYEIVLMQPTAD 60
 DB 1 MKLCSLAVLPVILFCEQHFVAFQSGOVLALPRTSRQVQLNLTYYEIVLMQPTAD 60
 QY 61 LIYKKQVHFVNASDVNDKKAHLNVSIGIPCSVLADVEDLIQOQISNDTVSPRASASY 120
 DB 61 LIYKKQVHFVNASDVNDKKAHLNVSIGIPCSVLADVEDLIQOQISNDTVSPRASASY 120
 QY 121 EGYHSLNEIYSWIEFTEHPDMLTKIHIGSFEEKYPLYLVKSGKEQTAKNAIWDGCI 180
 DB 121 EGYHSLNEIYSWIEFTEHPDMLTKIHIGSFEEKYPLYLVKSGKEQTAKNAIWDGCI 180
 QY 181 HAREWISPAFLMFIGH-----NRWRK 203
 DB 181 HAREWISPAFLMFIGHITQFYGIIGYTNLRLVDFYVMPVNVVDGYDSMKKNRMWRK 240
 QY 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSEYTCGLYPESEPEVKAASFRLRNI 263
 DB 241 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSEYTCGLYPESEPEVKAASFRLRNI 300
 QY 264 NQIKAYISMHSYQHIVFPYSTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGSGE 323
 DB 301 NQIKAYISMHSYQHIVFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSGE 360
 QY 324 TLYLAPGGDDWIVDGIKYSFT 346
 DB 361 TLYLAPGGDDWIVDGIKYSFT 383

RESULT 5
 US-08-696-139-2
 Sequence 2, Application US/08696139
 Patent No. 5672496
 GENERAL INFORMATION:
 APPLICANT: Fayerman, Jeffrey T.
 APPLICANT: Greenen, David P.
 APPLICANT: Hershtberger, Charles L.
 APPLICANT: Larson, Jeffrey L.
 APPLICANT: Sterner, Jane L.
 APPLICANT: Zhang, Haichao
 TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
 TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Eli Lilly and Company
 STREET: Lilly Corporate Center
 CITY: Indianapolis
 STATE: Indiana

Db 240 NONSTCIGTDLNRF-DVSDSSPNTNKPCLNVRGPAPESEKETKAVTFIRSHLSIK 298
Qy 268 AYISMHSYOHIVFPYSYTRSKSDHEELSLVASEAVRAIDTSKTRTHGSEETLYL 327
Db 299 AYTFHSYSOMLLIPGYTFKLPNHDLLKVARIAIDAL-STRETETRYIGPIASTYIK 357
Qy 328 APGGDDWITDGIKYSF 345
Db 358 TSGSLDMWYDGIKHTF 375

RESULT 9
US-08-430-787A-7

/ Sequence 7, Application US/08430787A
/ Patent No. 5593674
/ GENERAL INFORMATION:
/ APPLICANT: Drayna, Dennis T., Eaton, Dan L.
/ TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
/ NUMBER OF SEQUENCES: 8
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Genentech, Inc.
/ STREET: 460 Point San Bruno Blvd
/ CITY: South San Francisco
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94080
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patln (Genentech)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/430,787A
/ FILING DATE: 27-APR-1995
/ CLASSIFICATION: 514
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/277,540
/ FILING DATE: 19-JUL-1994
/ APPLICATION NUMBER: 08/167727
/ FILING DATE: 15-DEC-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/959944
/ FILING DATE: 14-OCT-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/649591
/ FILING DATE: 01-FEB-91
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hasak, Janet E.
/ REGISTRATION NUMBER: 28,616
/ REFERENCE/DOCKET NUMBER: 689D1C1D1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415/225-1896
/ TELEFAX: 415/952-9881
/ TELEX: 910/371-7168
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 417 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ US-08-430-787A-7

Query Match 32.9%, Score 628.5; DB 1; Length 417;
Best Local Similarity 37.0%, Pred. No. 9.5e-60;
Matches 140; Conservative 60; Mismatches 131; Indels 47; Gaps 8;

Qy 6 LAVLVPIVLFCEQHVAFQSGVLAALPRTSHQOVYQNLTTTYEIVLMQVYADLIYKK 65
Db 7 MAVITTLALAPVH--FDREKVFVKLQNEKHASVYLNKLDQSIDLDWYDPADIHDAVN 63
Qy 66 KOVHFVNASDVNDVKALNLSGICSVLADVELLQQOIS-NDTVSPRASASTYEYOH 124

Db 64 MTFVDFRSEKESQTIQSTLEQKHIEYELLHDLQEELEKOPFDKDELAGRHS---YAKYN 120
Qy 125 SLNEIYSWIEFIERPDMILTKIHISSEPKYPLYLYKVSGEQTKAKNAIMWDCGIHARE 184
Db 121 DMPKIVSWTEKMLEKHPHMSRIKISITVEDNPLYLKI-GKDGERRAIFMDCGIHARE 179
Qy 185 WISPAFLMEI-----GHN-----HMMRRNRSEF 207
Db 180 WISPAFCQMFVYQATKSYGKNKIMTKLDRMNFYLPVENVDCIYMSWTQDMMKRRNSR 239
Qy 208 YANNHCIGTDLNSNFVSKMCEGASSSCSEYCGLPSEPEYKAVASFLRRINQIK 267
Db 240 NONSTCIGTDLNRF-DVSDSSPNTNKPCLNVRGPAPESEKETKAVTNFIRSHLSIK 298
Qy 268 AYISMHSYOHIVFPYSYTRSKSDHEELSLVASEAVRAIDTSKTRTHGSEETLYL 327
Db 299 AYTFHSYSOMLLIPGYTFKLPNHDLLKVARIAIDAL-STRETETRYIGPIASTYIK 357
Qy 328 APGGDDWITDGIKYSF 345
Db 358 TSGSLDMWYDGIKHTF 375

RESULT 10
US-09-171-945-125

/ Sequence 125, Application US/09171945
/ Patent No. 6277599
/ GENERAL INFORMATION:
/ APPLICANT: Emery, Stephen
/ APPLICANT: Copley, Clive Graham
/ TITLE OF INVENTION: Monoclonal Antibody to CEa, Conjugates Comprising Said
/ TITLE OF INVENTION: Antibody, and Their Therapeutic Use In an Adept System
/ FILE REFERENCE: Monoclonal Antibody to CEa
/ CURRENT FILING DATE: 1998-10-29
/ PRIOR APPLICATION NUMBER: GB9703103.3
/ PRIOR FILING DATE: 1997-02-14
/ PRIOR APPLICATION NUMBER: GB9609405.7
/ PRIOR FILING DATE: 1996-05-04
/ PRIOR APPLICATION NUMBER: PCT/GB97/01165
/ NUMBER OF SEQ ID NOS: 131
/ SOFTWARE: Patentln Ver. 2.1
/ SEQ ID NO 125
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: humanized
/ US-09-171-945-125

Query Match 32.6%, Score 623; DB 4; Length 716;
Best Local Similarity 34.6%, Pred. No. 9.1e-59;
Matches 132; Conservative 72; Mismatches 134; Indels 44; Gaps 7;

Qy 3 LGSIAVLVPIVLFCEQHVFA-FQSGVLAALPRTSHQOVYQNLTTTYEIVLMQVYADLI 61
Db 1 MLALIVLVYVALASAHNGEHFEGEKYFRVNVEDENHNINIRLASTYQIDFWKPDVYQ 60
Qy 62 IVKKQVHFVNASDVNDVKALNLSGICSVLADVEDLIQQOISNDTVSPRASASTYE 121
Db 61 IKPHSTYDFRKADDTYVENVLQNELQYKVLISLNKNVVEAQFDSRV--RATGHSYE 117
Qy 122 QYHSLNEIYSWIEFIERPDMILTKIHISSEPKYPLYLYKVSGEQTKAKNAIMWDCGIH 181
Db 118 KYNKMETEAMTQOVATBNPALISRSVIGTTFEBRATYILKV-GKAGQNKPAIFMDCGFH 176
Qy 182 AREWISPAFLCMTF-----GH-----NRMMRN 204
Db 177 AREWISPAFCQMFVREAVRTYGREIQVTELLDKIDFYVLVPLNIDGIYIYWTJSRFRK 236

```

Oy 205 RSFYANNICIGDILNSNFVSKMCCEGASSSCSTYCGLYPESEPVEKVAASFILRRNTN 264
    ||: ||: |||| || || || || || || || || || || || || || || || || ||:
Db 237 RSTHGSSCIGTDPPRRNF-DAGWCIGTGRNCCDDETYGPAAESKETKALDFTRNKIS 295
Oy 265 OIKAYISMHSYQHIIVFPYSTYRSKSKDHEELSLVASEAVRAIDRTSKNTRYTHGHGSET 324
    ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
Db 296 SIKAVLTIHYSQOMIIVPYSTAYKRGENNABELMALAKATVKRL-ASLHGTXYTYGCAATT 354
Oy 325 LYLABGGDDMTYDGLGIRYSFT 346
    |: | || || || || || || || || || || || || || || || || || || || ||
Db 355 IYPSAGTSKDMAYVDGIRYSFT 376

RESULT 11
US-07-649-591B-6
: Sequence 6, Application US/07649591B
: Patent No. 5206161
: GENERAL INFORMATION:
: APPLICANT: Dennis Drayna and Daniel Eaton
: TITLE OF INVENTION: NO. 5206161el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/649,591B
: FILING DATE: 19910201
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 889
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/266-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 417 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
US-07-649-591B-6

Query Match 32.4%; Score 619.5; DB 1; Length 417;
Best Local Similarity 34.1%; Pred. No. 9e-59;
Matches 130; Conservative 76; Mismatches 128; Indels 47; Gaps 8;

Oy 6 LALVLPYILFCQHFVA--FOSGOYLAALPRTSQVOYOLWLTYYEILVLMQPYADLI 62
    : ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||:
Db 1 MRLILPVGILATTTAIAIPRFDEKVFYRVKPODERQADIINKLARTNELDFEYPGATHIV 60
Oy 63 YKKKQVHFVNASDVNKAHLNVSIGPQSVLLADVEDLIOOIS-NDTVSPRASASYE 121
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 AANMMVDRVSKESQAIOSALDOKMKMTEIILHDIQETIEKQFVKEDIPGRHS--YA 117
Oy 122 QYHSNLEYTWIEFTTERHPDMLTKIHIGSSPEKYPLVLYKSGKEQTAKNAIWDICGII 181
    ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||:
Db 118 KYNNEKEIVAWTEKMKMDKPEKVSRIKIGSTVEDNPVLYLKI-GEKNERRKAIIFDCGII 176
Oy 182 AREWISPAFCIMFTI-----GH-----NRMWRKN 204

```

```

Db      177 AREWSPAFCEQGFYVQAOKTVGRANKIMTRKLLDRMNFILPVFNVDGYIWSMTWKNNRMKRKN   | :         |||||
Oy      205 RSFYANNHCIGDLSNRYNSVKHWCCEGASSSCSETTCGLIPSEPEPKAAVSFLRRNTIN    264   || :||||| || : |
Db      237 RSKNONSNCICIDLRNP-NASWMSBIPMNDCADANRCSAPESEEKEKAAYNVFIRSHLN     295   || :||||| || : |
Oy      265 QIKATISMHSYSQHIVFPSTSYRSKSMDHEEELSVASEAARAVIDTSKTVRTTHGSGEST    324   || :||||| || : |
Db      296 EKUYVTITHSISQMILFFPGITSKLPPNHEDLAKYAKISTDVLT-STREYTRIITYGPDEST    354   || :||||| || : |
Oy          325 LYLAPGGDDWIYDLGIKYSF 345       :| | ||||| :|
Db          355 IYPISGSIDLWAYDI GKHTFE 375       :| | ||||| :|

RESULT 12
US-08-277-540-6
; Sequence 6, Application US/08277540
; Patent No. 5474901
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: NO. 3474901el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIAN TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/M$-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 669DICIDI1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO.: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-277-540-6

Query Match . 32.4%; Score 619.5; DB 1; Length 417;
Best Local Similarity 34.1%; Pred. No. 9e-59;
Matches 130; Conservative 76; Mismatches 128; Indels 47; Gaps 8;
```

```

QY 63 VKKQVHEFYVNASDVNDVNAHLNVSGIPCSVLLADVEDLIQOOIS-NDTVSPRASASYE 121
DB 61 AAMMMVDFRSEKESQALDQNMHYEILIHDLQEEIEKQFVKEDIPGRHS---YA 117
QY 122 QYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYVAKSGEQTAKNAIMWDCGIH 181
DB 118 KYNMKEIYAMTEKMDKPEMYSRIKIGSTVEDNPLYVAKI-GENNERKAI-FMDCGIH 176
QY 182 AREWISPAFCMLFI-----GH-----NRMWRKN 204
DB 177 AREWSPAFQCMFVYQATKTYGRNKMIMTKLLDRMNFYILPVENVDGYIWSWTKNRMWRKN 236
QY 205 RSFYANNHCIGTDLNSNEVSKHWCESGSSCSEYTGILYPESEPEVKAVASFLRRNIN 264
DB 237 RSKNONSCKICIGTLNRF-NASWNSIPNTNDPCADNYRGSAPESKEETKAVTNFRSHLN 295
QY 265 QIKAVISMHSYQIHFPPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGSGSET 324
DB 296 EIKVYITFHYSQMLLFPFGYTSKLPNNHEDLAKVAKIGTDVL-STREYETRYTIGPIEST 354
QY 325 LYIAPGGDDWYIDGIRYSF 345
DB 355 IYPIGSSLDWAYDIDGIRKHTF 375

```

RESULT 13

```

US-08-430-787A-6
; Sequence 6, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674e1 Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689DJC1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

US-08-430-787A-6

Query Match 32.4%; Score 619.5; DB 1; Length 417;
 Best Local Similarity 34.1%; Pred. No. 9e-59;
 Matches 130; Conservative 76; Mismatches 128; Indels 47; Gaps 8;

```

QY 6 LAVLPVLEFCQHVFA---FQSGVLAALPRTSRQVQVQLMTTYEIVLMQPTADLI 62
DB 1 MRLILPVLGLATTLAIPVRFREKVFYRKPODERQADILKPLATNELDFWYPCATHV 60
QY 63 VKKQVHEFYVNASDVNDVNAHLNVSGIPCSVLLADVEDLIQOOIS-NDTVSPRASASYE 121
DB 61 AAMMMVDFRSEKESQALDQNMHYEILIHDLQEEIEKQFVKEDIPGRHS---YA 117
QY 122 QYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYVAKSGEQTAKNAIMWDCGIH 181
DB 118 KYNMKEIYAMTEKMDKPEMYSRIKIGSTVEDNPLYVAKI-GENNERKAI-FMDCGIH 176
QY 182 AREWISPAFCMLFI-----GH-----NRMWRKN 204
DB 177 AREWSPAFQCMFVYQATKTYGRNKMIMTKLLDRMNFYILPVENVDGYIWSWTKNRMWRKN 236
QY 205 RSFYANNHCIGTDLNSNEVSKHWCESGSSCSEYTGILYPESEPEVKAVASFLRRNIN 264
DB 237 RSKNONSCKICIGTLNRF-NASWNSIPNTNDPCADNYRGSAPESKEETKAVTNFRSHLN 295
QY 265 QIKAVISMHSYQIHFPPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGSGSET 324
DB 296 EIKVYITFHYSQMLLFPFGYTSKLPNNHEDLAKVAKIGTDVL-STREYETRYTIGPIEST 354
QY 325 LYIAPGGDDWYIDGIRYSF 345
DB 355 IYPIGSSLDWAYDIDGIRKHTF 375

```

RESULT 14

```

US-08-782-760-6
; Sequence 6, Application US/08782760
; Patent No. 5948668
; GENERAL INFORMATION:
; APPLICANT: Hartman, Jacob
; APPLICANT: Fulga, Netta
; APPLICANT: Wendelovitch, Simona
; APPLICANT: Gorecki, Marian
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,760
; FILING DATE: 13-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/378,233
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0336/43847
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400

```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:38:53 ; Search time 65 Seconds

(without alignments)
615.178 Million cell updates/sec

Title: us-09-980-881-2

Perfect score: 1911
Sequence: 1 MKLCSLAVLPVILFCEQHV.....IKYFTSNPVEKILPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

1: A.Geneseq_032802:*
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1911	100.0	360	21	AA11457 Human brain carbox
2	1799.5	94.2	423	17	AA190293 Human plasma carbox
3	1799.5	94.2	423	18	AA190293 Human plasma carbox
4	1799.5	94.2	423	20	AA190293 Human plasma carbox
5	1794.5	93.9	423	14	AA190293 Human plasma carbox
6	1785	93.4	423	14	AA190293 Human plasma carbox
7	1325	69.3	423	21	AA190293 Human plasma carbox
8	803.5	42.0	211	21	AA190293 Human plasma carbox
9	640.5	33.5	404	16	AA190293 Human plasma carbox
10	639	33.4	415	17	AA190293 Human plasma carbox
11	639	33.4	415	19	AA190293 Human plasma carbox

12	631.5	33.0	424	17	AA190293 Human plasma carbox
13	629.5	32.9	424	22	AA190293 Human plasma carbox
14	623	32.6	716	18	AA190293 Human plasma carbox
15	604	31.6	307	17	AA190293 Human plasma carbox
16	604	31.6	437	22	AA190293 Human plasma carbox
17	576	30.1	306	16	AA190293 Human plasma carbox
18	576	30.1	306	22	AA190293 Human plasma carbox
19	563	29.5	354	22	AA190293 Human plasma carbox
20	563	29.5	354	22	AA190293 Human plasma carbox
21	563	29.5	354	22	AA190293 Human plasma carbox
22	563	29.5	374	22	AA190293 Human plasma carbox
23	563	29.5	374	22	AA190293 Human plasma carbox
24	561	29.4	349	17	AA190293 Human plasma carbox
25	561	29.4	349	17	AA190293 Human plasma carbox
26	554	29.0	349	17	AA190293 Human plasma carbox
27	554	29.0	349	17	AA190293 Human plasma carbox
28	553	28.9	349	17	AA190293 Human plasma carbox
29	553	28.9	349	17	AA190293 Human plasma carbox
30	550	28.8	349	18	AA190293 Human plasma carbox
31	548	28.7	349	18	AA190293 Human plasma carbox
32	548	28.7	349	18	AA190293 Human plasma carbox
33	547	28.6	349	18	AA190293 Human plasma carbox
34	546	28.6	349	18	AA190293 Human plasma carbox
35	546	28.6	349	18	AA190293 Human plasma carbox
36	545	28.5	349	18	AA190293 Human plasma carbox
37	545	28.5	349	18	AA190293 Human plasma carbox
38	545	28.5	349	18	AA190293 Human plasma carbox
39	545	28.5	349	18	AA190293 Human plasma carbox
40	543	28.4	349	18	AA190293 Human plasma carbox
41	543	28.4	349	18	AA190293 Human plasma carbox
42	543	28.4	613	19	AA190293 Human plasma carbox
43	542	28.4	349	18	AA190293 Human plasma carbox
44	542	28.4	349	18	AA190293 Human plasma carbox
45	540	28.3	349	18	AA190293 Human plasma carbox

ALIGNMENTS

RESULT 1	AA11457 standard; Protein: 360 AA.
ID	AA11457
AC	AA11457
XX	01-MAR-2001 (first entry)
DE	Human brain carboxypeptidase B protein.
XX	Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW	treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW	hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW	Down's syndrome; head trauma.
OS	Homo sapiens.
PN	WO20006717-A1.
XX	09-NOV-2000.
PD	01-MAY-2000; 2000WO-JP02878.
XX	30-APR-1999; 99JP-0125169.
PR	(MATS/) MATSUMOTO A.
PA	Matsumoto A;
XX	WPI; 2000-687534/67.
XX	N-PSDB; AAC81962.
DR	Human brain carboxypeptidase B isolated from the hippocampus useful for
XX	PT

PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Claim 1: Page 68-71; 84pp; Japanese.

CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimer's,
CC neurotropic, neuroprotective and hemostatic activity, and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening for promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.

XX Sequence 360 AA;

Query Match 100.0%; Score 1911; DB 21; Length 360;
Best Local Similarity 100.0%; Pred. No. 9,4e-186;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKICSLAVLPVLYLFCQHFVAFQSGOVLALPRTSROYVLOMLTTEYELVMQPYTAD 60
DB 1 MKICSLAVLPVLYLFCQHFVAFQSGVLAALPRTSRGVVLAALTTCTYELVWQPYTAD 60
OY 61 LLYKKQVHFVNASDVNVKAHLNVSGIRCSVLLADVEDLIQOQISNDVSPRASASY 120
DB 61 LLYKKQVHFVNASDVNVKAHLNVSGIRCSVLLADVEDLIQOQISNDVSPRASASY 120
OY 121 EGYHSLNEIYSWIEFTEHHPDLTKIHIGSSFEKYPVLYLVKSGKEQAKNAIWDGCI 180
DB 121 EGYHSLNEIYSWIEFTEHHPDLTKIHIGSSFEKYPVLYLVKSGKEQAKNAIWDGCI 180
OY 181 HAREWISPAFCIMFIGNHMRKNSFYANHCIGTDLNSFNFSKHCEGASSSCSCT 240
DB 181 HAREWISPAFCIMFIGNHMRKNSFYANHCIGTDLNSFNFSKHCEGASSSCSCT 240
OY 241 YGILPSEEPKAVASFIIRNLNGIKAYLSMYSQHVIFYSYLRKSKHEELSLVA 300
DB 241 YGILPSEEPKAVASFIIRNLNGIKAYLSMYSQHVIFYSYLRKSKHEELSLVA 300
OY 301 SEAVNAIDTKSKNTRYTHGSEETLYLAPGGDDMIYDLGIRYSEFTSNPVEKLLPLSLK 360
DB 301 SEAVNAIDTKSKNTRYTHGSEETLYLAPGGDDMIYDLGIRYSEFTSNPVEKLLPLSLK 360

RESULT 2

AAR90293 AAR90293 standard; Protein; 423 AA.

XX AAR90293;
XX
XX 12-Apr-1996 (first entry)

XX Human plasma carboxypeptidase B.

KW Plasma carboxypeptidase B: hpcpb; antibody; detection;
KW purification; plasminogen; affinity column.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Active-site 181 /note- "Residue suspected of being involved in
FT catalytic activity of enzyme."

FT Binding-site 183 /note- "Residue suspected of being part of the
FT binding site of enzyme."

FT Binding-site 256..257 /note- "Residues suspected of being part of the
FT binding site of enzyme."

FT Active-site 310

FT /note- "Residue suspected of being involved in
FT catalytic activity of enzyme."
FT Binding-site 312 /note- "Residue suspected of being part of the
FT binding site of enzyme."
FT Active-site 369 /note- "Residue suspected of being involved in
FT catalytic activity of enzyme."
FT Binding-site 370..371 /note- "Residues suspected of being part of the
FT binding site of enzyme."
FT Binding-site 394 /note- "Residue suspected of being part of the
FT binding site of enzyme."

XX US5474301-A.

XX 12-DEC-1995.

XX 01-FEB-1991; 91US-0649591.

XX 01-FEB-1991; 91US-0649591.

XX 14-OCT-1992; 92US-0959944.

XX 15-DEC-1993; 93US-0167727.

XX 19-JUL-1994; 94US-0277540.

XX (GETH) GENENTECH INC.

XX Dreyne DT, Eaton DL.

XX WPI: 1996-039508/04.

XX N-PSDB; AAT11671.

XX Antibody to human plasma carboxypeptidase B - useful for detecting
XX and purifying hpcpb for use in treating clotting disorders e.g.
XX haemophilia A

XX Disclosure; Figure 4; 40pp; English.

XX An antibody which specifically binds human plasma carboxypeptidase B
XX (hpcpb) and does not cross react with other carboxypeptidases is
XX useful for the detection of hpcpb in vitro. The antibody is also
XX used for purifying hpcpb from a sample. Purification comprises
XX passing a sample thought to contain hpcpb over either a column to
XX which antibody has been bound, or a plasminogen affinity column,
XX eluting the column and then recovering the fraction containing the
XX hpcpb.

XX Sequence 423 AA;

Query Match 94.2%; Score 1799.5; DB 17; Length 423;
Best Local Similarity 89.6%; Pred. No. 3.7e-176;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

OY 1 MKICSLAVLPVLYLFCQHFVAFQSGOVLALPRTSROYVLOMLTTEYELVMQPYTAD 60
DB 1 MKICSLAVLPVLYLFCQHFVAFQSGVLAALPRTSRGVVLAALTTCTYELVWQPYTAD 60

OY 61 LLYKKQVHFVNASDVNVKAHLNVSGIRCSVLLADVEDLIQOQISNDVSPRASASY 120
DB 61 LLYKKQVHFVNASDVNVKAHLNVSGIRCSVLLADVEDLIQOQISNDVSPRASASY 120

OY 121 EGYHSLNEIYSWIEFTEHHPDLTKIHIGSSFEKYPVLYLVKSGKEQAKNAIWDGCI 180
DB 121 EGYHSLNEIYSWIEFTEHHPDLTKIHIGSSFEKYPVLYLVKSGKEQAKNAIWDGCI 180

OY 181 HAREWISPAFCIMFIGNHMRKNSFYANHCIGTDLNSFNFSKHCEGASSSCSCT 240
DB 181 HAREWISPAFCIMFIGNHMRKNSFYANHCIGTDLNSFNFSKHCEGASSSCSCT 240

OY 204 NRSFYANHCIGTDLNSFNFSKHCEGASSSCSCTSEYVCLYSESEPVAAVSFLRRNT 263
DB 204 NRSFYANHCIGTDLNSFNFSKHCEGASSSCSCTSEYVCLYSESEPVAAVSFLRRNT 263

DB 241 nrsfyannhcgldlnrnfskhwceagssscsetyqlypesepevkavasflrrnl 300

QY 264 NOIKAVISMHSYSQHIYFPYSTRSKSDHELSLVAEAVRAIDKTSKNTRYTHGSGE 323

DB 301 ngikayismhsyqghlyfpystrskskdheelslvaseavratektskntrythgsgse 360

QY 324 TLYLAPGGGDDWLYDLGIRKISFT 346

DB 361 tlylapgggddwlydlgikysft 383

RESULT 3

AAW14733

ID AAW14733 standard; Protein; 423 AA.

XX AAW14733;

AC AAW14733;

XX 08-MAY-1997 (first entry)

DE Human plasma carboxypeptidase B.

XX Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;

KM Plasma; plasmalogen.

XX Homo sapiens.

OS

XX Key

FH Location/Qualifiers

FT 1..22

FT /note= "Signal peptide"

FT Protein

FT 23..423

FT /note= "Mature PCPB"

FT 114

FT Cleavage-site

FT /note= "Potential clip site for activation of PCPB as a carboxypeptidase"

FT 181

FT Active-site

FT /note= "Expected to be involved in catalytic activity"

FT 183

FT Binding-site

FT /note= "Expected to be involved in substrate binding"

FT 184

FT Active-site

FT /note= "Expected to be involved in catalytic activity"

FT 256

FT Binding-site

FT /note= "Expected to be involved in substrate binding"

FT 257

FT Binding-site

FT /note= "Expected to be involved in substrate binding"

FT 310

FT Active-site

FT /note= "Expected to be involved in catalytic activity"

FT 312

FT Binding-site

FT /note= "Expected to be involved in substrate binding"

FT 369

FT Active-site

FT /note= "Expected to be involved in catalytic activity"

FT 370

FT Binding-site

FT /note= "Expected to be involved in substrate binding, thought to determine specificity of PCPB as a carboxypeptidase B"

FT 371

FT Binding-site

FT /note= "Expected to be involved in substrate binding"

FT 394

FT Binding-site

FT /note= "Expected to be involved in substrate binding"

XX US5593674-A.

PN 14-JAN-1997.

XX 14-JAN-1997.

PD

XX 01-FEB-1991;

PF 91US-0649591.

XX 01-FEB-1991;

XX 91US-0649591.

PR 14-OCT-1992;

PR 92US-0959944.

PR 15-DEC-1993;

PR 93US-0167727.

PR 19-JUL-1994;

PR 94US-0277540.

PR 27-APR-1995;

XX 95US-0430787.

XX

PA (GETH) GENENTECH INC.

PI Drayna DT, Eaton DL;

XX

DR WPI: 1997-099413/09.

DR N-PSDB; AAT62846.

XX

PT Using human plasma carboxypeptidase B in blood coagulation - 1s

PT functionally related to carboxypeptidase A and pancreas

PT carboxypeptidase B

XX

PS Example 2; Column 37-42; 39pp; English.

XX

CC This sequence represents human plasma carboxypeptidase B (PCPB) which

CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.

CC PCPB may be used therapeutically in haemostatic regulation. PCPB is

CC purified from human plasma or by transformed cell culture by

CC extraction using plasmalogen bound to a solid phase.

XX

SQ Sequence 423 AA;

Query Match

Best Local Similarity 89.6%; Score 1799.5; DB 18; Length 423;

Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

QY 1 MKLCSLAVLPVLPFCEQHVPFQSGOYLALPRTSRQVOVLNLTYYEIVLMQPTAD 60

DB 1 mklsclavlpvlpfceqhvpfqsqgylaalprrtsrqvqlnlttyelvwqpvfad 60

QY 61 LIYKKQVHFVNASDNDNVKAHLNVSGIPCSVLADVEDLIOOISNDTVSPRASA 120

DB 61 liykkqvhfvnasdndnvkahlnvsgipcsvlladvedlqqisndtvsprrasay 120

QY 121 EOYHSLNEIYSWIEFITERHHPDMLTKIHGSSFEKYPYLVKVGKEQTAKNAIMICGI 180

DB 121 eqynslneiyswiefiterhpdmltkihgssfekypylvkvgkeqtaknalwidqgi 180

QY 181 HAREWISPAFLWFTGH-----NRMAR 203

DB 181 harewispaflwftghlftgfygltlnllrvdlyvmpvrvndgydyswkknmrwrk 240

QY 204 NRSFYANNHCIGTDLNLSFVSKHWCCEGASSSCSEYTCGYPSEPEVKAVASFLRRNI 263

DB 241 nrsfyannhcgldlnrnfskhwceagssscsetyqlypesepevkavasflrrnl 300

QY 264 NOIKAVISMHSYSQHIYFPYSTRSKSDHELSLVAEAVRAIDKTSKNTRYTHGSGE 323

DB 301 ngikayismhsyqghlyfpystrskskdheelslvaseavratektskntrythgsgse 360

QY 324 TLYLAPGGGDDWLYDLGIRKISFT 346

DB 361 tlylapgggddwlydlgikysft 383

RESULT 4

AAW92270

ID AAW92270 standard; Protein; 423 AA.

XX AAW92270;

AC AAW92270;

XX 28-APR-1999 (first entry)

DT

XX 28-APR-1999 (first entry)

XX

XX Human plasma carboxypeptidase B (PCPB) thr147.

XX

XX Human plasma carboxypeptidase B; PCPB; human; hPCPBthr147;

KM polymorphism detection; thrombotic disease.

XX

OS Homo sapiens.

XX

XX Key

FH Location/Qualifiers

FT 1..22

FT Peptide

FT /note= "signal peptide"

FT Protein

FT 23..423

FT /note= "mature PCPBthr147"

CC between the deduced preproPCPB amino acid sequence and pro-human
 CC mast cell carboxypeptidase A and between preproPCPB and prepro-rat
 CC carboxypeptidase B. Human PCPB has the same substrate binding sites
 CC as, and shares 6 cysteine residues which form 3 intramolecular
 CC disulphide bonds with. bovine and rat carboxypeptidase B. The
 CC presence of the same amino acid (asp at position 348 of the mature
 CC protein) at the region in carboxypeptidases that determines
 CC substrate specificity, suggests that PCPB represents a plasma-
 CC derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
 CC tPA of plasminogen to plasmin in the presence of fibrinogen.

XX Sequence 423 AA:

Query Match 93.9%; Score 1794.5; DB 14; Length 423;
 Best Local Similarity 89.3%; Pred. No. 1.2e-175;
 Matches 342; Conservative 1; Mismatches 3; Indels 37; Gaps 1;

QY 1 MKICSLAVIVPILVPCQOHFAFGSGOVLAAIPRTSRNOGVQNLMTTYEIVLMQPTAD 60
 DB 1 MKICSLAVIVPILVPCQOHFAFGSGOVLAAIPRTSRQVILNLTLYEIVLWQPTAD 60
 QY 61 LTVKKQVHFVNASVDNVAHNLNVSGLPCSVLADVEDLIQOQISNDTVSPRASASY 120
 DB 61 LTVKKQVHFVNASVDNVAHNLNVSGLPCSVLADVEDLIQQISNDTVSPRASASY 120
 QY 121 EQYHSNIEYSWIEFTEERRHDMLEKIHGSSFEKYPVLVYLVKSGKEQTAKNAIWDGCI 180
 DB 121 eqyhsneiyswiefteerrhdmlekihgsfekypylvlvksykeqtaknaivdgcgl 180
 QY 181 HAREWISPAFCIMFIGH-----NRMRK 203
 DB 181 harewispfcilwfighligfyglgqytnlllvdfvmpvvnvvdgydyswknrmrk 240
 QY 204 NRSFYANNHCIGDNLNSNFYSKHMCEGASSSCSEYCGLYPESPEVAVASFLEARNI 263
 DB 241 nrsfyannhncigdlrnntaskhwccegaassscsetyqglypesepevavasfltrnl 300
 QY 264 NOIKAYISMHSYQSHIVFPYSYTRRSKSDHEELSLVASEAVRAIDKTSKTRTTHGSGS 323
 DB 301 ngikayismhsyqshivfypsyrtrskskdheelslvaseavratelktsktrtylhgsgs 360
 QY 324 TLYLAPGGDDWITLDGIRKYSFT 346
 DB 361 tlylapggddwilydlgikysft 383

RESULT 6

AAB11458
 ID AAB11458 standard; Protein; 338 AA.

XX AAB11458;

DT 01-MAR-2001 (first entry)

DE Human brain carboxypeptidase B protein SEQ ID NO 3.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KW cerebroprotective; antialzheimers; neurotrophic; neuroprotective;
 KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KW Down's syndrome; head trauma.

XX Homo sapiens.

XX WO200066717-A1.

PD 09-NOV-2000.

PF 01-MAY-2000; 2000WO-JP02878.

PR 30-APR-1999; 99JP-0125169.

XX

PA (MATS/) MATSUMOTO A.

PI Matsumoto A;

DR WPI; 2000-687534/67.

XX Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -

PS Disclosure; Page 71-73; 84pp; Japanese.

CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimers,
 CC neurotrophic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.

XX Sequence 338 AA:

Query Match 93.4%; Score 1785; DB 21; Length 338;
 Best Local Similarity 99.1%; Pred. No. 7.9e-175;
 Matches 335; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 23 FQSGOVLAALPRTSROVOVQLNLTTEYEVLMQVNTADLYKKQVHFVNASVDNVKA 82
 DB 1 fgsqgvlaalprtsrqvqlnltteyevlwgqvntadlyvkkqgvhfvnasvdnvka 60
 QY 83 HLNVSGLPCSVLADVEDLIQOQISNDTVSPRASASYEQYHSNIEYSWIEFTEERRHPD 142
 DB 61 hlnvsglpcsvladvdeliqqisndtvspreasasyeqyhsneiyswiefteerrhpd 120
 QY 143 MLTKIHGSSFEKYPVLVYLVKSGKEQTAKNAIWDGCIHAREWISPAFCIMFIGHNRMR 202
 DB 121 mltkihgsfekypylvlvksykeqtaknaivdgcgharewispfcilwfighnmrmr 180
 QY 203 KNRSEYANNHCIGDNLNSNFYSKHMCEGASSSCSEYCGLYPESPEVAVASFLEARNI 262
 DB 181 knrsfyannhncigdlrnntaskhwccegaassscsetyqglypesepevavasfltrn 240
 QY 263 INQIKAYISMHSYQSHIVFPYSYTRRSKSDHEELSLVASEAVRAIDKTSKTRTTHGSGS 322
 DB 241 inqikayismhsyqshivfypsyrtrskskdheelslvaseavratelktsktrtylhgsgs 300
 QY 323 ETVYLABGGDDWITLDGIRKYSFTSNPPVEKLPPLSLK 360
 DB 301 etvylapggddwilydlgikysftsnppveklplslk 338

RESULT 7

AAB11459
 ID AAB11459 standard; Protein; 246 AA.

XX AAB11459;

DT 01-MAR-2001 (first entry)

DE Human brain carboxypeptidase B protein SEQ ID NO 4.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KW cerebroprotective; antialzheimers; neurotrophic; neuroprotective;
 KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KW Down's syndrome; head trauma.

XX Homo sapiens.

XX WO200066717-A1.

XX

XX 09-NOV-2000.
 PD
 XX
 PF 01-MAY-2000; 2000WO-JP02878.
 XX
 PR 30-APR-1999; 99JP-0125169.
 XX
 PA (MATSU/) MATSUMOTO A.
 XX
 PI Matsumoto A;
 XX
 DR WPI: 2000-687534/67.
 XX
 PT Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 XX
 PS Disclosure: Page 74-75; 84pp; Japanese.
 XX
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimers,
 CC neurotropic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 CC
 SQ Sequence 246 AA;
 Query Match 69.3%; Score 1325; DB 21; Length 246;
 Best Local Similarity 98.8%; Pred. No. 1e-127;
 Matches 243; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 115 ASASYVEOHSINETYSMETTERHPDLTKIHGSPREKYLVLVYSGEOTAKNAI 174
 DB 1 aasayveghelneleyweftehpdmnlkthysstekyplylvkvsgeqaknal 60
 QY 175 WDCGITHAREWISPAFCIMFICGHNMRKNSFYANNHCIGTDLNSNFYSKHCCEGASS 234
 DB 61 wdcgitharewispafciwffighnmwrknsfyannhcgldlnrnfaskhwccegass 120
 QY 235 SSCSEYCGIVPESEPEYKAVASFLRRNINQKATISMSYSQHIYFPRYSYTRSKSDHE 294
 DB 121 sscseycgilypesepevkavasflrrlnlqkayismysqhlvfpystrskskdhe 180
 QY 295 ELSLVASEAVRAIDKTSKNTRYTHGSEETLYLAPGGDDWIYDLGIRYSFTSNPPEVKL 354
 DB 181 elslvaseavralektsknttrythgsetlylapggddwilydlgikysftsnppvekl 240
 QY 355 LPLSLK 360
 DB 241 lplslk 246
 RESULT 8
 AAB58129
 ID AAB58129 standard; Protein; 211 AA.
 AC AAB58129;
 XX
 XX 14-MAR-2001 (first entry)
 DE
 XX
 XX Lung cancer associated polypeptide sequence SEQ ID 467.
 XX
 XX Human; lung cancer associated protein; neuroprotective; cytostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnery;
 KW gastrointestinal; nephrotropic; antinefective; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease.
 XX

OS Homo sapiens.
 XX
 PN WO200055180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PA (ROSE/) ROSEN C A.
 XX
 PI Ruben SM;
 XX
 DR WPI: 2000-587514/55.
 DR N-PSDB; AAF18005.
 DR
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 11; Page 944-945; 1425pp; English.
 XX
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective, cytostatic, cardioactive,
 CC immunomodulatory, muscular active, general; vulnery; gastrointestinal
 CC general; nephrotropic; antinefective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterization of the polynucleotide and protein
 CC sequences.
 SQ Sequence 211 AA;
 Query Match 42.0%; Score 803.5; DB 21; Length 211;
 Best Local Similarity 88.8%; Pred. No. 3.8e-74;
 Matches 151; Conservative 2; Mismatches 8; Indels 9; Gaps 2;
 QY 185 WISPAFCIMFT-----GHNRRMRKNSFYANNHCIGTDLNSNFYSKHCCEGASSS 236
 DB 3 wis-mfcrlmmwmywsknmwrknsfyannhcgldlnrnfaskhwccegasss 61
 QY 237 CSEYCGIVPESEPEYKAVASFLRRNINQKATISMSYSQHIYFPRYSYTRSKSDHEEL 296
 DB 62 cseycgilypesepevkavasflrrlnlqkayismysqhlvfpystrskskdheel 121
 QY 297 SLVASEAVRAIDKTSKNTRYTHGSEETLYLAPGGDDWIYDLGIRYSFTSNPPEVKL 346
 DB 122 slvaseavralektsknttrythgsetlylapggddwilydlgikysft 171
 RESULT 9
 AAR75131
 ID AAR75131 standard; Protein; 404 AA.
 AC AAR75131;
 XX
 XX 13-MAR-1996 (first entry)
 DE
 XX
 XX Porcine Tyr-His-Met Procarboxypeptidase B.
 XX

RESULT 11

```
AAW74476
ID   AAW74476 standard; Protein: 415 AA.
XX
XX   AAW74476;
AC
XX
XX   20-NOV-1998 (first entry)
DE
XX
XX   Human pancreatic carboxypeptidase protein sequence.
KW
XX   Human; pancreatic carboxypeptidase B; insulin; protein sequencing;
OS
XX   Homo sapiens.
PN
XX   WO9835988-A1.
PD
XX   20-AUG-1998.
PF
XX   10-FEB-1998;    98WO-GB00415.
PR
XX   29-OCT-1997;     97GH-0022727.
PR
XX   14-FEB-1997;     97GB-0003104.
PR
XX   18-OCT-1997;      97GB-0022003.
XX
XX   (ZENE ) ZENEPHA (ZENE ) ZENEPHA LTD.
XX
XX   Edge MD:
XX
XX   WPI: 1998-467168/40.
XX
XX   N-PSDB: AAV41795.
PT
XX
XX   New modified pro-domain of carboxy-peptidase B - enhances expression
PT
XX   of co-expressed proteins for production of recombinant
PT
XX   carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT
XX   enzyme producting therapy
XX
XX   Example 1; Page 54-55; 83pp; English.
```

The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a modified pro-domain of CPB on a separate gene to enhance recombinant production. This process can be used to produce recombinant CPB in eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in insulin production and protein sequencing, while its fusions with

CC antibodies are useful in antibody-directed enzyme prodrug therapy. The
CC modified pro-domain provide Increased yields of recombinant CPB, possibly
CC by protecting the C-terminus against enzymatic degradation or by
CC increasing intracellular trafficking.

50 Sequence 415 AA;

Query Match	33.48;	Score 639;	DB 19;	Length 415;
Best Local Similarity	35.68;	Pred. No. 9.7e-57;		
Matches 135;	Conservative 69;	Mismatches 131;	Indels 44;	Gaps 7;

[illegible]

XX		RESULT_12
XX		AAM06175
ID		AAM06175 standard; Protein; 424 AA.
AC		
AA		AAM06175;
DT		17-FEB-1997 (first entry)
XX		
DE		ProHCPB with pelB leader sequence.
XX		
KW		ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW		mustard-ribonucleotide; antibody directed enzyme prodng therapy;
KW		anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW		reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
Fd		F(ab') ₂ ; PelB; leader; human carboxypeptidase B.
XX		
OS		Synthetic.
XX		
Key		Location/Qualifiers
Peptide		1..22
PFT		/label= pelB_leader_sequence
PFT		23..117
PFT		/label= pro_sequence
Protein		118..424
PFT		/label= mature_HCPB
PFT		
XX		
PN		W09620011-A1.
XX		
PD		04-JUL-1996.
XX		
PF		21-DEC-1995;
XX		95WO-CB02991.

```

PR 16-AUG-1995; 95GB-0016810.
PR 23-DEC-1994; 94GB-0026192.
XX
XX (ZENE ) ZENECA LTD.
XX
PI Blakey DC, Boyle FT, Davies DR, Eggelte HJ, Heaton DM;
PI Hennam JF, Hennesquin LFA, Marsham PR, Rabin BR, Slater AM;
PI Tarragona-Fiol A, Taylorson CJ;
XX
XX WPI; 1996-321650/32.
XX
XX N-PSDB; AAT42506.
XX
XX Two component system for anti-tumour therapy - comprising targeting
XX anti-neoplastic produg
XX
XX Reference Example 18; Page 140-142; 182pp; English.
XX
XX A two-component system for anti-tumour therapy comprises a targeting
XX moiety linked to a mutated enzyme which can transform an anti-neoplastic
XX produg. The system is based on antibody directed enzyme produg therapy
XX (ADEPT) using non-naturally occurring mutant forms of host enzymes,
XX pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
XX Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
XX be used. The present sequence is that of proHCPB with a Pelt leader
XX sequence encoded by pIC11738.
XX
XX Sequence 424 AA;

Query Match 33.0%; Score 631.5; DB 17; Length 424;
Best Local Similarity 35.7%; Pred. No. 6e-56;
Matches 129; Conservative 68; Mismatches 121; Indels 43; Gaps 6;

QY 23 FOSQVLAALPRTSROYQONTTTEIYVMPVZADLYKKQVHEFNASDVNDYKA 82
DB 29 fegekvtrrvvedenhnlirelastqtdtqkpsdvtqlkphstvdfrvkaedltvven 88
QY 83 HLNVSGIPCVLADVEDLIQOQISNDTVSPRASASYEYQHSNLETYSWIEFTEHRPD 142
DB 89 vlkqnelqykvllsnlnvveaqdsrv---ratghayekynkellavwqatempa 145
QY 143 MLTKIHGSSPEKRYPLVYLVKSGKEQTAKNAIWDGCIHAREWISPAFLMFT----- 195
DB 146 llsrsvlgtfegrailkv-gkaagqkallndcgfhawelspafcgwfvreaavrtly 204
QY 196 -----GH-----NRMWRKRSFYANNHCIGTDLSNFYSK 225
DB 205 grelqvteilldkldfyilpvlndgylytwktsfrwrtktsbhgscigtdpnrf-daa 263
QY 226 HMCESGASSSSCSEYTCGLYPESEPEVKAVASFLRRINQIKAYISMHSYSQHIYFPYSY 285
DB 264 gwceigsarncpdcetycgpaeseketkaldfirnklsikayllthysyqmmllypysy 323
QY 286 TRSKSKHHELSIVASRAVAIDKTSKNTRTYTHGSETLYLAGGGDDWYIDGIKYSF 345
DB 324 aylkgemnaelnaakatlvel-aslngtkylygpatcltlypaagsddwagdqlrystf 382
QY 346 T 346
DB 383 t 383

```

```

KW secreted metallocarboxypeptidase; immunomodulatory;
KW antiparkinsonian; nootropic; anticonvulsant; neuroprotective;
KW osteopathic; antiarthritic; cerebroprotective; vasotropic; vulnery;
KW antibacterial; antiviral; antifungal; antiinflammatory; antidiabetic;
KW antineumatic; cytostatic; antiallergic; haemostatic; infection;
KW cancer.
XX
XX Bothops Jataraca.
XX
XX WO200157265-A1.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US03783.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 29-SEP-2000; 2000US-0676135.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Boyle BJ, Mize NK, Arterburn MC, Yeung G, Tang YT, Liu C;
XX Drmanac RT, Wang M, Chen L, Yang Y;
XX
XX WPI; 2001-465705/50.
XX
XX Novel metallocarboxy peptidase-like polypeptide is useful in treating
XX neurodegenerative, immune, etc. disorders
XX
XX Example 3; Fig 1; 146pp; English.
XX
XX The invention relates to an isolated polypeptide comprising an amino
XX acid sequence of 374, 41, 14, 15, 41, 22, 15, 27, 14, 9, 18, 20, 15,
XX 354, 165, 374, 354, or 41 amino acids, fully defined in the
XX specification, or its translated protein coding portion. Mature protein
XX coding portion, extracellular portion or active domain. The polypeptide
XX is a human secreted metallocarboxypeptidase-like polypeptide. It
XX is useful for treating mammalian diseases including:
XX neurodegenerative diseases such as Parkinson's or Alzheimer's diseases,
XX osteoporosis, osteoarthritis, disorders linked to reduced tissue
XX growth, Huntington's disease, amyotrophic lateral sclerosis, stroke,
XX reperfusion, wound healing, infectious diseases, immune and autoimmune
XX disorders such as severe combined immunodeficiency, rheumatoid
XX arthritis, Guillain-Barre syndrome, insulin dependent diabetes
XX mellitus, graft-versus-host disease and allergy, cancer, and
XX haemophilia. The polypeptide is also useful in medical imaging, as an
XX immunogen to generate antibodies, as a nutritional source or
XX supplement, and as a contraceptive. The present sequence is
XX a carboxypeptidase homologue protein of Bothops Jataraca.
XX
XX Sequence 324 AA;

Query Match 32.9%; Score 629.5; DB 22; Length 324;
Best Local Similarity 41.7%; Pred. No. 6.2e-56;
Matches 125; Conservative 53; Mismatches 79; Indels 43; Gaps 5;

QY 84 LNVSGIPCVLADVEDLIQOQISNDTVSPRASASYEYQHSNLETYSWIEFTEHRPD 143
DB 2 lqsgslngyellldnlgavldtqlqdhna---rtagynyekynswekidtdianemsl 58
QY 144 LTKIHGSSPEKRYPLVYLVKSGKEQTAKNAIWDGCIHAREWISPAFLMFT----- 195
DB 59 vsrlqigtftgrmpmlkv-gkpgvnxkalfidcgfharewelspafcgwfvreaavrtly 117
QY 196 -----GNNRMWRKRSFYANNHCIGTDLSNFYSK 226
DB 118 ketlmtqllnkldfyilpvlndgylyvswksmrwrtktrsvnaagstcigtprnrf-daa 176
QY 227 WCEEGASSSSCSEYTCGLYPESEPEVKAVASFLRRINQIKAYISMHSYSQHIYFPYSY 286
DB 177 wcvsgasrnpesetcygskpeseketkaldfirnrslsqayllthysyqmmllypysyt 236

```

OY 287 RSKSKDHEELSLVASEAVRAIDKSKNRTYHGHSEFLYLPAGGDDWYDLGKXSEF 346
 Db 237 ydltsnmknklnslakelrei-kvligteylygpgaatlypaagsddwagdglykyaft 295

RESULT 14

AAW41414
 ID AAW41414 standard; Protein; 716 AA.

AC AAW41414;

DT 02-JUN-1998 (first entry)

DE PreproHCPB-linker-humanised 806.077 VH.

KW Anti-CEA antibody; carcinoembryonic antigen; 806.077 Ab; cancer therapy;
 cancer diagnosis; complementarity determining region.

OS Synthetic.

PN W09742329-A1.

PD 13-NOV-1997.

PF 29-APR-1997; 97WO-GB01165.

PR 14-FEB-1997; 97GB-0003103.
 04-MAY-1996; 96GB-0009405.

PA (ZENE) ZENECA LTD.

PI Copley CG, Edge MD, Emery SC;

DR WPI: 1997-558987/51.

DR N-PSDB; AAV17331.

PT Anti-carcinoembryonic antigen antibody 806.077 Ab - used for
 diagnosis and therapy of cancer

PS Example 75; Page 184-187; 208pp; English.

CC This sequence is preproHCPB-linker-humanised 807.077 VH, which is
 an example of the antibody of the invention. The antibody is an anti-CEA
 (carcinoembryonic antigen) antibody (preferably 806.077 Ab). Host cells
 or transgenic organisms transformed with DNA encoding the antibody, are
 used to make the antibody or conjugate. The conjugate is used in a
 medicament suitable for intravenous administration. The conjugate can be
 used for cancer therapy, selectively killing tumour cells. The antibody
 can be used for in vivo or in vitro diagnosis of cancer.

CC Sequence 716 AA;

Query Match 32.6%; Score 623; DB 18; Length 716;

Best Local Similarity 34.6%; Pred. No. 1.1e-54;

Matches 132; Conservative 72; Mismatches 134; Indels 44; Gaps 7;

OY 3 LCSLAVLYPIVIFCQHFA-FQSCQVLAALPRTSRQOVQVQLNTTTEIYIMQPVTA 61
 Db 1 mllllylvtvalasahghegekefrvndenhlirclastcldqldfwkpsvtyg 60

OY 62 IVKKQVFEFVNASDVNDVKAHLNVSIGIPSVLADVEDLIQQQISNDTVSPRASAYTE 121

Db 61 lkphtvfrkaedvtvenvlnkqnelgykvlslntlnveagdsrv--ratghsye 117

OY 122 QYHSLNETYSWIEFTTERHPDMLTKIHGSSFEKYPYLVYKVGSKEDQAKNAIWDGCIH 181

Db 118 kynkwtelavtqvatentpalrsvlygttegralyllkv-gkagqnpkalfmdcgfh 176

OY 182 AREWISPAFLCMFI-----GH-----NDMNRKN 204
 Db 177 arewispafqwfvravrtlygrtelvclldkldfyvlpvlnldgylytwcksrfrwkt 236

OY 205 RSFYANNHCIGDLDNSNFVSKHWCCEGASSSCSEFTYCGLYPSEPEVNAVASFLRRNIN 264
 Db 237 rslhtgssclgtprnfn-dagwcelgaarncpdcelycpaaseketaladfnkls 295

OY 265 QIKATYSMHSYSOHIVFPYSTRSKDHEELSLVASEAVRAIDKSKNRTYHGHSEF 324

Db 296 slkaylthnsyqumlypysayklgenaelnalakavkel-ashbtktylypggatt 354

OY 325 LYLAPGGDDWYDLGKXSEF 346

Db 355 lypsagtskdwagdglystf 376

RESULT 15

AAW00602
 ID AAW00602 standard; Protein; 307 AA.

AC AAW00602;

DT 09-NOV-1996 (first entry)

DE Rat carboxypeptidase B mature enzyme.

KW Carboxypeptidase-B; pro-enzyme; protease; insulin.

OS Rattus sp.

PN W09623064-A1.

PD 01-AUG-1996.

PF 25-JAN-1996; 96WO-US00995.

PR 25-JAN-1995; 95US-0378233.

PA (BIOT-) BIOTECHNOLOGY GEN CORP.

PI Fulga N, Gorecki M, Hartman J, Mendelovitch S;

DR WPI: 1996-362688/36.

DR N-PSDB; AAT35760.

PT Purified active recombinant carboxypeptidase B prodn. - by
 expressing DNA encoding the pro-enzyme, folding and enzymatic
 cleavage to give active form, used e.g. for insulin prodn.

PS Example 1; Page 37-38; 49pp; English.

CC The amino acid sequence of the rat mature carboxypeptidase-B
 (CPB) (AAW00602) was deduced from a cDNA sequence (see also AAT35760)
 obtd. by PCR amplification of Sprague-Dawley rat pancreas cDNA.
 CC The sequence of the activation peptide (AAW00601) of pro-CPB has
 also been deduced. Pro-CPB (pref. with a C290S mutation) can be
 expressed in host cells, e.g. Escherichia coli, and subsequently
 recovered, refolded and cleaved with trypsin to yield the active
 enzyme. CPB produced this way is cheaper than porcine pancreatic
 enzyme, and is free of other proteases.

CC Sequence 307 AA;

Query Match 31.6%; Score 604; DB 17; Length 307;

Best Local Similarity 44.6%; Pred. No. 2.4e-53;

Matches 120; Conservative 43; Mismatches 66; Indels 40; Gaps 5;

OY 115 ASASYEDYHSLNETYSWIEFTTERHPDMLTKIHGSSFEKYPYLVYKVGSKEDQAKNAI 174

Db 1 asghsytklynmetelavtqvatndpdlvtqsvlygttegrmnyvlki-gktrpnkpa 59

OY 175 WIDCGIHAREWISPAFLCMFI-----GH----- 197
 Db 60 fidegfharewispafqwfvravrtlygrtelvclldkldfyvlpvlnldgylytwck 119

QY 198 NRMWRKRSPFYANNHCIGTDLNSNFVSKHWCCEGASSSCSEYCGLYPESEPEVKAVAS 257
Db 120 drrmrkLrslmagssclgvprnrf-nagwcevgasrpscsetycgpapesekckalad 178
QY 258 FLRRNINQIKAYISMHSYSOHIVEPYSTRSKSKDHEELSLVASEAVRAIDKTSKNTRYT 317
Db 179 flrrnisltkayltlinsyqgmlypyeydyklpenyeelnalvgaakel-atlhgtkylt 237
QY 318 HGHGSETLYLAPGGDDWYLDGIKYSFT 346
Db 238 yppgatllypaagsddwsydgikysft 266

Search completed: September 18, 2002, 16:45:11
Job time: 378 sec

This Page Blank (uspto)

PT haemophilia A
 XX
 PS Disclosure: Figure 4; 40pp; English.
 CC
 CC An antibody which specifically binds human plasma carboxypeptidase B
 CC (hPCPB) and does not cross react with other carboxypeptidases is
 CC useful for the detection of hPCPB in vitro. The antibody is also
 CC used for purifying hPCPB from a sample. Purification comprises
 CC passing a sample thought to contain hPCPB over either a column to
 CC which antibody has been bound, or a plasmaogen affinity column,
 CC eluting the column and then recovering the fraction containing the
 CC hPCPB.
 CC
 XX Sequence 423 AA:

alignment_scores:
 Quality: 1885.00 Length: 423
 Ratio: 5.122 Gaps: 2
 Percent Similarity: 86.998 Percent Identity: 86.525

alignment_block:
 US-09-980-881-1 x AAR90293 ..

Align seg 1/1 to: AAR90293 from: 1 to: 423

```

18 ATGAAGCTTTCGACGCTTGCAGTCTGTACCATTTGTTCTTCTGTGA 67
1 MetLysLeuCysserLeuAlaValLeuValProIleValLeuPheCysgl 17
68 GCACGATGCTTCGCGTTCACAGAGTGGCCAAAGTTCTAGCTCTCTCTTA 117
17 uGlnHisValAlaPheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProA 34
118 GAACCTTAGGCAAGTCAAGTTCTACAGAACTTACTACAAATATATAG 167
34 rGlnSerAlaGlnValGlnValLeuGlnAsnLeuThrThrThrTyrGln 50
168 ATGTGCTCTGGCAGCGGTAAACAGTACCTTATTTGAAAGAAAACA 217
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysgl 67
218 AGTGCATTTTGTGTAATGATCTGATGTGGAATGTAAGCCCATTT 267
67 nValHisPhePheValAlaAsnAlaSerValAlaAsnValLysAlaHisL 84
268 TAAATGTGAGGAGATTCATGATGAGTGTCTGTGGCAGAGCTGAGAT 317
84 euAsnValSerGlyIleProCysSerValIleuLeuAlaAspValGlnAsp 100
318 CTATATGACAGCAGATTTCCACAGACAGTACGCCCCGAGCCTCCGC 367
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaGlnAlaSerAl 117
368 ATCTACTATGACAGATGATCTACTCAATGAATGAAATCTATTCTGGATAG 417
117 aserTyrTyrGlnGlnIleTyrHisSerLeuAsnGlnIleTyrSerTrpIleG 134
418 AATTATTAAGTACAGAGCATCCGATATGTTTACAAAAAATCCACATTTGA 467
134 luePheIleThrGlnAlaGlnHisProAspMetLeuThrLysIleHisIleLeu 150
468 TCCATCATTTGAGAGTACCCACTCTATGTTTAAAGTTTCTGAGAAAAGA 517
151 SerSerPheGlnLysTyrProLeuTyrValLeuLysValSerGlyLysgl 167
518 ACAAAACGCCCCAAAATGCCATATGATGATGGAATTCATGCCAGAG 567
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyLleHisAlaArg 184
568 AATGATCTCTCCGCTTCTGCTTGATGATAGGCCAT..... 608
184 lueTrpIleSerProAlaPheCysLeuThrPheIleGlyHisIleThrGln 200
  
```

```

608 ..... 608
201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspH 217
608 ..... 608
217 eTyrValMetProValValAsnValAspGlyTyrAspTyrSerTrpLysL 234
609 ..AATCGAATGTGGAGAAAGAACCGTTCTTCTATGACAAATCATGTCG 656
234 ysAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCys 250
657 ATCGAACAGACCTGAATATAGCACTTTGTCTCCAAACACTGGTGTAGGA 706
251 IleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGlnG 267
707 AGTGCATTCACAGTTCCTCATGCTCGGAAACCTACCTGTGAGATTTATCTG 756
267 uGlyAlaSerSerSerSerCysSerGlnThrTyrCysGlyLeuTyrProG 284
757 AGTCAGAACAGAAAGTGAAGCAGTGGCTAGTCTTGTGAGAAATATC 806
284 lueSerGlnProGlnValLysAlaValAlaSerPheLeuArgArgAsnIle 300
807 AACGATTTAAAGCATACATGACATGATCATCTCCAGCATATAGT 856
301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVa 317
857 GTTTCATATTCCTTATACAGCAAGTAAAGCAAAAGCATGAGAACGT 906
317 lPheTrpTyrSerTyrThrArgSerLysSerLysAspHisGlnGlnLeuS 334
907 CTCTAGTACGACAGTACAGTGTGCTATTTGACAAACTAGTAAAT 956
334 eTLeuValAlaSerGlnValAlaArgAlaIleGlnLysThrSerLysAsn 350
957 ACCAGGTATACAGATGCCATGGCTCAGAAACCTTATACCTAGCTCTGG 1006
351 ThrArgTyrThrHisGlyHisGlySerGlnThrLeuTyrLeuAlaProG 367
1007 AGGTGGGACCATTTGGATCTATGATTTGGCATCAAAATTCGTT..... 1051
367 yGlyGlyAspAspTrpIleTyrAspLeuGlyLleLysTyrSerPheThrI 384
1052 .....TAC 1054
384 leGlnLeuAlaArgAspThrGlyThrTyrGlyPheLeuLeuProGlnArgTyr 400
1055 ATCAAAACCCACCTGTAGAGAGCTTTGCGCGTGTCTTAAATAGCTTG 1104
401 IleLysProThrCysArgGlnAlaPheAlaAlaValSerLysIleAlaTr 417
1105 GCATGTCATTTAGCAATGTT 1123
417 PHisValIleAlaArgAsnVal 423

seq_name: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1197.DAT:AAW14733
seq_documentation_block:
ID AAW14733 standard; Protein: 423 AA.
XX
AC AAW14733:
XX
DT 08-MAY-1997 (first entry)
XX
XX Human plasma carboxypeptidase B.
XX
XX Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
XX
XX Homo sapiens.
XX
  
```

Key	Location/Qualifiers
Peptide	1..22
Protein	/note= "Signal peptide"
Cleavage-site	23..423
	/note= "Mature PCPB"
Active-site	114
	/note= "Potential clip site for activation of PCPB as a carboxypeptidase"
Binding-site	181
	/note= "Expected to be involved in catalytic activity"
Active-site	183
	/note= "Expected to be involved in substrate binding"
Binding-site	184
	/note= "Expected to be involved in catalytic activity"
Binding-site	256
	/note= "Expected to be involved in substrate binding"
Active-site	257
	/note= "Expected to be involved in substrate binding"
Binding-site	310
	/note= "Expected to be involved in catalytic activity"
Active-site	312
	/note= "Expected to be involved in substrate binding"
Binding-site	369
	/note= "Expected to be involved in catalytic activity"
Active-site	370
	/note= "Expected to be involved in substrate binding, thought to determine specificity of PCPB as a carboxypeptidase B"
Binding-site	371
	/note= "Expected to be involved in substrate binding"
Binding-site	394
	/note= "Expected to be involved in substrate binding"
US593674-A.	
14-JAN-1997.	
01-FEB-1991;	91US-0649591.
01-FEB-1991;	91US-0649591.
14-OCT-1992;	92US-0649591.
15-DEC-1993;	93US-0167727.
19-JUL-1994;	94US-0227540.
27-APR-1995;	95US-0430787.
(GETH) GENENTECH INC.	
Drayna DT, Eaton DL;	
WPI: 1997-099413/09.	
N-PSDB; AAT62846.	
Using human plasma carboxypeptidase B in blood coagulation - is functionally related to carboxypeptidase A and pancreas carboxypeptidase B	
Example 2: Column 37-42: 39pp: English.	
This sequence represents human plasma carboxypeptidase B (PCPB) which has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD. PCPB may be used therapeutically in haemostatic regulation. PCPB is purified from human plasma or by transformed cell culture by extraction using plasminogen bound to a solid phase.	
Sequence 423 AA:	

Alignment_scores:
 Quality: 1885.00
 Ratio: 5.122
 Percent Similarity: 86.998
 Percent Identity: 86.525

Alignment_block:

```

US-09-980-881-1 x AAW14733 ..
Align seg 1/1 to: AAW14733 from: 1 to: 423

18 ATGAGCTTGGACGCTTGGACGCTTGTACCATTTCTTCTGTGA 67
1 MetLysLeuCySserLeuAlaValLeuValProIleValLeuPheCySgl 17
68 GCACGATGCTCGCTCCAGAGTGGCAAGTTCGTGCTCTTCCTTA 117
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaIleAlaLeuPro 34
118 GAACCTCTGAGCAAGTCAAGTCTTACAGAACTTACTACACATATGAG 167
34 rGThrSerArgGlnValGlnValLeuGlnAsnLeuThrTrpTrpGlu 50
168 ATTGTTCTCTGGCAGCCGGTACAGCTGACCTTATGTGAGAAAAACA 217
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGln 67
218 AGTCATTTTGTAAATGATCATCTGATGTCGACAAATGGAAGCCCAT 267
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
268 TAAATGAGCGGAATTCATGCAAGTGTCTTGTGCGCAGACGTGAAGAT 317
84 euAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
318 CTATTCACACAGCAATTTCCACGACGACGACCCCGACCTCCGC 367
101 LeuIleGlnGlnIleSerAsnAspThrValSerProAlaSerAl 117
368 ATGCTATATGACAGTATCATCTACTAAATGAATCTTCTTGATAG 417
117 aserGlyTrpGluGlnIleTrpHisSerLeuAsnGlnIleTrpSerTrpIleG 134
418 AATTTATACTGAGAGCATCTGATATGCTTCAAAAATCCATATGGA 467
134 IuPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
468 TCCTCATTTGAGAGTACCACTATGTTTAAAGTTTCTGGAAAGA 517
151 SerSerPheGlnLysTrpProLeuTrpValLeuLysValSerGlyLysGln 167
518 ACAACAGCCAAAATGCCATATGATGATGATGATGATGATGATGATG 567
167 uGlnThrAlaLysAsnAlaIleThrIleAspCySglIleHisAlaArg 184
568 AATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 608
184 IuTrpIleSerProAlaPheCySLeuTrpPheIleGlyHisIleThrGln 200
608 ..... 608
201 pheTyGlyIleIleGlyGlnTrpAsnLeuLeuArgLeuValAspPh 217
608 ..... 608
217 eTyValMetProValValAsnValAspGlyTrpAspTrpLysL 234
609 .. AATGGAATGGGAGAAAGAACCGTTCTTCTATGCGACACATATGCG 656
234 yAsnArgMetTrpArgLysAsnArgSerPheTyAlaAsnHisCyS 250
657 ATCGGAACAGACCTGATAGCAACTTGTCTCAACACACTGGTGAGGA 706
251 IleglyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCySglu 267
707 AGGTGATCCAGTCTCTCATGCTCGGAAACCTAGTGTGACATTATCTG 756
267 uGlyAlaSerSerSerCySserGlnThrTyCySglLeuTyProG 284
757 ACTCAGAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAGAAATATC 806

```

```

|||||
284 lusergluproglnvalalalaserpheuleuargasnile 300
807 AACCAATTAAACATCATCATCATCTACTCCAGCATATAGT 856
301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVa 317
857 GTTTCATATTCCTATACAGAAATAAAGCAAGACCATGAGACTGT 906
317 lPheProTyrSerTyrThrArgSerTyrSerLysAspHisGlnIleuLeuS 334
907 CTCTAGAGCAGTGAAGAGCTGCTATGTAACAACTAGCAAAAT 956
334 erleuValAlaSerGlnAlaValAlaGlnIleGlnTyrSerLysAsn 350
957 ACCAGATATACATGCGCATGCGTCAGAAACCTTATACCTAGCTCTGG 1006
351 ThrArgTyrThrHisGlnHisGlnSerGlnThrLeuTyrIleAlaProG1 367
1007 AGGTGGGAGCATTTGATCATGATTTGGGCATCAAAATATTCGTT.... 1051
367 YGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThrI 384
1052 .....TAC 1054
384 legIuleuArgAspTrpGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
1055 ATCAAAACCCACCTGTAGAGAGCTTTGGCGCTGCTCTAAATAGCTTG 1104
401 lIleLysProThrCysArgGlnAlaPheAlaAlaValSerLysIleAlaTr 417
1105 GCATGTCATTAGGAATGTT 1123
417 PHISValIleArgAsnVal 423

seq_name: /studs1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:AAW92270

seq_documentation_block:
ID AAW92270 standard; Protein; 423 AA.
AC
AAW92270;
XX
XX 28-APR-1999 (first entry)
XX
XX Human plasma carboxypeptidase B (PCPB) thr147.
XX
XX Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
XX
XX polymorphism detection; thrombotic disease.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 1..22
XX FT /note="signal peptide"
XX FT 23..423
XX FT Protein /note="mature PCPBthr147"
XX
XX WO9855645-A1.
XX
XX PD 10-DEC-1998.
XX
XX PF 02-JUN-1998; 98WO-EP03244.
XX
XX PR 03-JUN-1997; 97US-0869057.
XX
XX PA (SCHD) SCHERING AG.
XX
XX PI Morser MJ, Nagashima M:
XX
XX WPI; 1999-045800/04.
XX
XX DR N-PSDB; AAV74302.
XX
XX PT Detecting new polymorphism of human plasma carboxypeptidase B -

```

```

PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
XX PS Disclosure; Page 25-26; 35pp; English.
XX
XX CC This sequence is the human plasma carboxypeptidase B (PCPB) mutant
XX CC hPCPBthr147. The invention relates to a method for determining the
XX CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
XX CC comprises obtaining a prepared tissue or blood sample and determining the
XX CC presence of DNA coding for naturally occurring polymorphs of the protein
XX CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
XX CC respectively). Determination of the relative distribution of the PCPB
XX CC polymorphs in a patient's blood by genetic or protein analysis by the
XX CC methods is useful to determine the risk of thrombotic disease in humans.
XX CC Such assessments may be made by accumulating information concerning the
XX CC relative distribution of the different polymorphs within the general
XX CC population compared with populations known to be at risk and establishing
XX CC a PCPB polymorph profile for at-risk patients.
XX
XX SQ Sequence 423 AA;

alignment_scores:
Quality: 1885.00 Length: 423
Ratio: 5.122 Gaps: 2
Percent Similarity: 86.998 Percent Identity: 86.525

alignment_block:
us-09-980-881-1 x AAW92270 ..

Align seq 1/1 to: AAW92270 from: 1 to: 423

18 ATGAAGCTTTCGAGCCTTGAGTCCTTGACCAATGTTCTCTGCTGGA 67
1 MetLysLeuCysSerLeuAlaValLeuValProIleValIleuPheCysG1 17
68 GCAGCATGCTCTTCGCGTTCAGAGTGGCCAAATTCTAGCTGCTTCTGA 117
17 uGlnHisValPheAlaPheGlnSerGlyGlnValIleuAlaIleuProA 34
118 GAACCTCTAGGCAAGTTCAATGTTCTACAGAACTTCTACACAACTATGAG 167
34 rGlnHisSerArgGlnValGlnValIleuGlnAsnLeuThrThrTyrGln 50
168 ATTGCTCTGCGCAGCCGTAACAGCTGACCTATTGGAAGAAAGAAACA 217
51 lIleValIleuTrpGlnProValThrAlaIleuIleValLysLysG1 67
218 AGTCATTTTCTTGTAAATGCATCTGATGTCAGCAATGTGAAGCCCAT 267
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
268 TAAATGTGAGCGGAATTCATGACAGTGTCTGTCGGCAGAGTGGAGAT 317
84 euAsnValSerGlyIleProCysSerValIleuLeuAlaAspValGlnuSp 100
318 CTATATCAACAGAGATTTCCACAGACAGCTCAGCCCGAGCCTCGCG 367
101 leuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaIleSerAl 117
368 ATGCTACTATGAGACAGTATCACTACCTAAATGAATCTATCTTGATGAG 417
117 aserTyrTyrGlnGlnIleTyrHisSerLeuAsnGlnIleTyrSerTrpIleG 134
418 AATTATACAGAGAGAGCATCTGATATGCTTACAAAAATCCACATATGGA 467
134 lIupheIleThrGlnArgHisProAspMetLeuThrLysIleHisIleLelY 150
468 TCTTCATTTGAGAGTACCACTCATGTTTAAAGTTCTGGAAGAAAGA 517
151 SerSerPheGlnLysTyrProLeuTyrValIleuLysValSerGlyLysG1 167

```

```

518 ACAACGCCAATAATGCGATATGATGACTGTGGAATCCATGCCAGAG 567
167 UGLTHRALALYSANALAIETRPILLESAPCYGLYLEHISALARG 184
568 AATGATCTCTCTGCTTGTGCTTGTGCTCATAGGCCAT..... 608
184 LUTRPILLESERPROIALPHECYSLEUTRPHEILEGLYHISILETHGLN 200
608 ..... 608
201 PHELYRGLYLEILEGLYGLNTYRTHRASNLEULEUARGLEUVALASPRH 217
608 ..... 608
217 ELYRVALMETPROVALVALASNAVALASPGLYTYRASPTYSERTRPYSL 234
609 .AATCGAATGTGAGAGAAAGACCGTTCTTTCATGCGAACATCATTTGC 656
234 YASANAARGMETTRPARGLYASANARGSERPHELYRALASANAHISCY 250
657 ATCGGAACAGACTGAATACCAACTTTGTCTCAACACATGCTGAGAGA 706
251 ILEGLYTHRASNLEUASNAARGSNPHEALASERTYSHSTRPCYSLUGL 267
707 AGGTGATCCAGTTCCTCATGCTGCGAAACCTACTGCTGACTTATCTG 756
267 UGLYALASERSESRSEYSEYSEYSEYSEYSEYSEYSEYSEYSEYSEY 284
757 AGTCAGAACAGAAAGTGAAGCAGTGGCTAGTTCTTGGAACAAATATATC 806
284 LUSERGLUPROGLUVALVALALASERPHELEUARGASNAILE 300
807 AACCAAGTTTAAAGCATACATCAGCATCTCATCTCCACAGCATATAGT 856
301 ASNGINLEYSALATYRILESERMECHISERTYSEGINHISILEVA 317
857 GTTCCATATTCCTATACAGAGTAAGTAAGCAAAAGACAGAGAACTGT 906
317 LPHEPROTYSETYRTHRARGSERLYSERLYSASPHISGLULUENS 334
907 CTCTAGTACCACTGAAGCAGTTCGTGCTATGTGACAAACTAGTAAAT 956
334 ERLEUVALALASERGLUVALALARGALALEGLULYSTHRSERYASNA 350
957 ACCAGGTATACATGCGCATGGCTCAGAAACCTTATACCTACCTCTGG 1006
351 THARGTYRTHRISGLYHISGLYSEYSEYSEYSEYSEYSEYSEYSEY 367
1007 AGGTGGGAGCATGATGATGATGATGATGATGATGATGATGATGATG 1051
367 YGLYGLYASAPSPTRPILERYRASPLEUGLYILEYSTYRSEYRPHETHT 384
1052 .....TMC 1054
384 LEGLULEUARGASPRHNGLYTHRGLYPHELEULEUPROGLUARGTYG 400
1055 ATCAACCCACCTGTAGAGAGCTTTTGGCGCTGCTCTAAATACCTTG 1104
401 ILEYSPTHTCYSGARGLUVALPHEALALVALISERTYSHILEALTR 417
1105 GCATGTCTATTAGATGT 1123
417 PHISVALILEARGASNAVAL 423

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:AA1993

seq_documentation_block: ID: AA1993 standard; Protein: 423 AA.

AC: AA1993; DT: 24-AUG-1993 (first entry)

xx

```

DE Human plasma carboxypeptidase B.
XX
XX PCPB, tissue plasminogen activator inhibitor; t-PA inhibitor.
KM
XX Homo sapiens.
OS
XX Key
XX Peptide
XX Location/Qualifiers
XX 1..22
XX /label= signal_peptide
XX /note= "putative site of cleavage by trypsin to
XX activate PCPB as a carboxypeptidase"
XX 181
XX /note= "expected to be involved in catalytic
XX activity"
XX 183
XX /note= "expected to be involved in substrate
XX binding"
XX 184
XX /note= "expected to be involved in catalytic
XX activity"
XX 256..257
XX /note= "expected to be involved in substrate
XX binding"
XX 310
XX /note= "expected to be involved in catalytic
XX activity"
XX 312
XX /note= "expected to be involved in substrate
XX binding"
XX 369
XX /note= "expected to be involved in catalytic
XX activity"
XX 370..371
XX /note= "expected to be involved in substrate
XX binding; Asp 370 is the residue which
XX determines specificity of the PCPB as
XX a carboxypeptidase B"
XX 394
XX /note= "expected to be involved in substrate
XX binding"
XX
XX US5206161-A.
XX
XX 27-APR-1993.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI. 1993-151724/18.
XX N-PSDB; AA041001.
XX
XX New human plasma carboxypeptidase B - used as haemostatic
XX regulator for clotting blood, partic. for treating blood clotting
XX disorders, e.g. haemophilia
XX
XX Claim 2; Fig 4; 40pp; English.
XX
XX This amino acid sequence was deduced from a clone isolated from a
XX human liver cDNA library. There is about 40% sequence identity
XX between the deduced preproPCPB amino acid sequence and pro-human
XX mast cell carboxypeptidase A and between preproPCPB and prepro-rat
XX as, and shares 6 cysteine residues which form 3 intramolecular
XX disulphide bonds with, bovine and rat carboxypeptidase B. The
XX presence of the same amino acid (Asp at position 348 of the mature
XX protein) at the region in carboxypeptidases that determines
XX substrate specificity, suggests that PCPB represents a plasma-

```


XX (MATS/) MATSUMOTO A.
 XX
 XX Matsumoto A;
 XX
 XX WPI; 2000-687534/67.
 XX
 PT Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 XX
 PS Disclosure; Page 71-73; 84pp; Japanese.
 XX
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity, and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 CC
 SQ Sequence 338 AA;

alignment_scores:
 Quality: 1785.00 Length: 338
 Ratio: 5.297 Gaps: 0
 Percent Similarity: 99.704 Percent Identity: 99.112

alignment_block:

US-09-980-881-1 x AAB11458 ..

Align seg 1/1 to: AAB11458 from: 1 to: 338

84 TTCGAGAGTGGCCAGTTCAGTCTCTAGAACCTTAGGCAAGT 133
 1 PheGlnSerIleValLeuAlaLeuProArgThrSerArgGlnVal 17
 134 TCAGATTCTACAGATCTTACTACACATATGAGATTGTTCTGCGAGC 183
 17 IGIValIleuGlnAsnLeuThrThrThrArgIleValIleuLeuTrpIle 34
 184 CGGTACAGCTGACCTTATTTGTAAGAAAAACAGCTCCATTTTGTGA 233
 34 rovalIThrAlaAspLeuIleValIleValIleValIleValIleVal 50
 234 AATGATCTGATGTGCAATGTGAAGCCCATTTAAATGAGCGGAAT 283
 51 AsnAlaSerIleValAspAsnValIleValIleValIleValIleVal 67
 284 TCCATCAGTGTCTGTGGCAGAGCTGGAAGATCTTATTCACACAGCA 333
 67 eProCysSerValIleuLeuAlaAspValGluAspLeuIleGlnGlnI 84
 334 TTTCGACGACAGTCAAGTCAAGCCCGGAGCTCCGCACTGACTTGAACG 383
 84 IeSerAsnAspIleValSerProArgAlaSerAlaSerIleValIleVal 100
 384 TATCACTCACTAAATGAATCTTATTTGATGAGATTATTAACAGAG 433
 101 TyrHisSerLeuAsnGluIleValIleValIleValIleValIleVal 117
 434 GCATCTGTATGCTTACAAAAATCCACATGTGATCCTCATTTGAGAGT 483
 117 gHisProAspMetIleuThrIleHisIleGlySerSerPheGluIle 134
 484 ACCGACTCATATGTTTAAAGTTCTGGAAGAAAGAAACAGCAAAAT 533
 134 yTrProIeuValIleuValSerGlyIleValIleValIleValIleVal 150
 534 GCCATATGATGTGACTGTGAAATCCATGCCAGAGATGATCTCTCTCG 583

151 AlaIleTrpIleAspCysGlyIleIleIleAlaArgIleUtrPheSerProAl 167
 584 TTTCGCTGTGGTTCATAGGCGCATATCGAATGTGGAAAGAACCGTT 633
 167 aPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArgIleGlyAsnArg 184
 634 CTTCCTATGCGAACATCATGTGATCCGAAACAGACAGCTGAATAGCACTT 683
 184 eRheTrpIleAlaAsnAsnIleCysIleGlyThrAspLeuAsnAlaGlnPhe 200
 684 GTTCGCAAAACACTGTGTGAGGAAGTGCATCCAGTTCTCATGCTCGGA 733
 201 AlaSerIleHisTrpCysGluGluGlyAlaSerSerSerSerSerSerG 217
 734 AACCTAGTGGACTTATATCTGTCAGAACAGAAAGTGAAGCAAGTGG 783
 217 UTrpTrpCysGlyLeuTrpProGluSerGluProGluValIleValAla 234
 784 CTAGTTCTTGAGAGAAATATCAACCCAGATTAAAGCATACATCAGCATG 833
 234 IAspPheLeuArgArgAsnIleAsnGlnIleValIleValIleValIle 250
 834 CATTCATACCTCCAGCATATAGTGTTCATATTCATATACAGAAATA 883
 251 HisSerTrpSerGlnHisIleValIlePheProTrpSerTrpThrArgSer 267
 884 AACCAAGACCATGAGGAGTGTCTGTATAGCCAGTGAAGCAAGTTCGTG 933
 267 SSerIleAspHisGluGluIleuSerIleValIleValIleValIleVal 284
 934 CTATTCGAAAACTAGTAAATATACAGATATACAGATGCGCATGCTCA 983
 284 IAlleGlyIleTrpSerIleAsnThrArgTrpThrHisGlyHisGlySer 300
 984 GAAACCTTATACCTAGCTCTGAGAGTGGGAGCATTTGATGATTT 1033
 301 GluThrLeuTrpLeuAlaProGlyGlyAspAspTrpIleTrpAspIle 317
 1034 GGGCATCAATATTCGTTTACATCAAAACCACTGTAGAGAACCTTTGC 1083
 317 uGlyIleIleValTrpSerPheThrSerAsnProProValGluValLeu 334
 1084 CGCTGTCTCTAATAA 1097
 334 roLeuSerIleuLys 338
 seq_name: /SID51/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.AAB11459
 seq_documentation_block:
 ID AAB11459 standard; Protein; 246 AA.
 XX
 XX AAB11459;
 XX
 XX
 XX 01-MAR-2001 (first entry)
 XX
 DE Human brain carboxypeptidase B protein SEQ ID NO. 4.
 XX
 XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KW cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
 KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KW Down's syndrome; head trauma.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200066717-A1.
 PN
 XX
 XX 09-NOV-2000.
 PD
 XX
 XX 01-MAY-2000; 2000WO-JP02878.
 PF
 XX
 XX 30-APR-1999; 99JP-0125169.
 PR

XX (MATS/) MATSUMOTO A.
 XX
 PI Matsumoto A;
 XX
 DR WPI: 2000-687534/67.
 XX
 PT Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 XX
 PS Disclosure: Page 74-75; 84pp; Japanese.
 XX
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 XX
 SQ Sequence 246 AA:

alignment_scores: Length: 246
 Quality: 1325.00 Gaps: 0
 Ratio: 5.408
 Percent Similarity: 99.593 Percent Identity: 98.780

alignment_block:

US-09-980-881-1 x AAB11459 ..

Align seq 1/1 to: AAB11459 from: 1 to: 246

360 GCCCTCCGACGCTGCTATGAGACGATACACGACATTAATGAAATCTATTC 409
 1 AlserlaSerlyrtyrGluGlnItyrHisSerleuasnGlnIleTyse 17
 410 TTGGATGAAATTTTAACTGAGAGGACGCGATGATGCTTCAAAAATCC 459
 17 rTPrIleGluPheIleThrGluAlaGHisProAspMetIleuThrIysIleH 34
 460 ACATTTGGATCCTCATTTGAGAAAGTACCACTGTATGTTTAAAGTTTCT 509
 34 IsIleGlySerSerPheGluIuIySerProleuTyValIleuIyValSer 50
 510 GGAAGAAACAAACAGCCAAAATGCCATATGATGATGACTGTGGAATGCA 559
 51 GlyIySGluGlnThrAlaIyAsnAlaIleTrpIleAspCysGlyIleH 67
 560 TGGCAGAGAAATGATCTCTCCTGCTTCTGCTTGGTTCATAGGCCATA 609
 67 sAlaArgGluTrpIleSerProAlaPheCysIleuTrpPheIleuIyHSA 84
 610 ATCGAATGTGGAGAAAGACCGTTCTTTCTATGCGAACAATCATTTGCATC 659
 84 snArgMetTrpArgIyAsnArgSerPheTyAlaAsnAsnHisCysIle 100
 660 GGAACAAACCTGGAATAGCACTTTGTCTCCAAACACTGGGTGAGGAAG 709
 101 GlIyThrspIeuAsnArgAsnPheAlaSerTyHisIstPrCysGluGluG 117
 710 TGCATTCAGTTCTCATGCTCGGAACCTACTGTGACTTATATCTGAGT 759
 117 yAlaSerSerSerSerSerGluTrpIyTyCysGlyIleuTyTrpGluS 134
 760 CAGAAGCAGAAAGTGAAGGACAGCTGCTAGTTTCTTGAGAAAGAAATATCAAC 809
 134 ergIuPProGluValIyAlaValAlaSerPheIleuArgArgAsnIleAsn 150
 810 CAGATTAAAGCATCATCATGATGATTCATCTCCAGCATATATAGTGT 859

151 GlnIleIySAlaTyrlIleSerMetHisSerTySerGlnHisIleValPh 167
 860 TCCATATTCCTATACAGAGTAAGCAAGACCATGAGCAACTGTCTC 909
 167 eProTySerTyThrArgSerTySerIyAspHisIleuGluIeuSerL 184
 910 TAGTAGCCAGTAGAAGACGTTGCTGTATTGACAAATACTAGTAAATAC 959
 184 euValAlaSerGluAlaValAlaArgAlaIleGluIyThrSerIyAsnThr 200
 960 AGGTTATACATGGCCATGGCTCAGAAACCTTATACCTAGTCTCGGAGG 1009
 201 ArgTyThrHisGlyHisGlySerGluThrIeuTyIleuAlaProGlyG 217
 1010 TGGGACGACGATGATCTATGATTTGGCGATCAAAATATCGTTTACATCA 1059
 217 yGlyAspAspTrpIleTyThrAspIleuGlyIleIyTySerPheThr 234
 1060 ACCGACCTGTAGAGAAAGCTTTGCCGCTGTCTCTTAA 1097
 224 snProProValGluIyIuIySerProIeuSerIeuIyS 246

seq_name: /STD1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA58129

seq_documentation_block:

ID AAB58129 standard; Protein; 211 AA.

AC AAB58129;

DT 14-MAR-2001 (first entry)

DE Lung cancer associated polypeptide sequence SEQ ID 467.

XX

KW Human: lung cancer associated protein; neuroprotective; cyrostatic;

KW gastrointestinal; immunomodulatory; muscular active; vulnetary;

KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;

XX Homo sapiens.

XX WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000MO-US05918.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Ruben SM;

XX WPI: 2000-587514/55.

XX N-PSDB; AAF18005.

XX Claim 11; Page 944-945; 1425pp; English.

CC Polynucleotide sequences AAF1982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytosolic; cardioactive;
 CC immunomodulatory; muscular active general; vulnetary; gastrointestinal
 CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer.

Percent Similarity: 64.678 Percent Identity: 33.890

alignment_block:

US-09-980-881-1 x AAM06172 ..

Align seg 1/1 to: AAM06172 from: 1 to: 415

```

33 CTTCGACGCTCTGACCATGTTCTCTCTGAGACAGCATGCTTCGCG 82
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGly 18
83 G... TTCCAGGTGCGCAAGTTCAGTCTGCTCTCCAGAACCTTGAGC 129
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 UNHSPHGLUGLYGLULYSVALPHEARGVALASNVALGLULSPGLUNSH 35
130 AAGTTCAGTCTACAGATCTTACTACACATATGATGATGCTCTGG 179
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 IStIeAnIleIleArgIleuAlaSerThrThrGlnIleAspPheTrp 51
180 CAGCCGCTAACGCTGACCTTATTGTGAAGAAAAACAGATCTATTTT 229
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 LysProAspSerValThrGlnIleLysProHisSerThrValAspPheAr 68
230 TGTAAATGATCAGTGTGCGCATGGAAGCCCATTTAATGTGAGCG 279
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 GValLysAlaGluAspThrValThrValGluAsnValLeuLysGlnAsnG 85
280 GAATTCATGACGATGCTGCTGCGCAGACGTGAAGATCTTATTCACAG 329
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 IuLeuGlnIlyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101
330 CAGATTTCCAAAGACAGACAGTCCGCGAGCTCCGATGCTACTATGA 379
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 GlnPheAspSerArgVal..... ArgAlaThrGlnHisSerLysGln 115
380 ACAGTATCAGCTACTAAATGATCTATCTTGATAGATTTTAACTG 429
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 ULysTrAsnLysTrpGlnThrIleGluAlaTrpThrGlnIleValAlaTr 132
430 AGAGCATTCCTGATATGCTTACAAAAATCCATGATGATCTCATTTAG 479
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 hrgIuAsnProAlaLeuIleSerArgSerValIleGlyThrThrPheGln 148
480 AAGTACCACCTATGTTTAAAGTTTCTGAAAGAAACAAACAGCCAA 529
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
149 GlyArgAlaIleTyrLeuLeuLysVal... GlyLysAlaGlyGlnAsnLys 164
530 AAATGCCATATGATGCTGATGCAATCCATGCCAGAGATGATCTTC 579
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 sProAlaIlePheMetAspCysGlyPheHisAlaArgGlnTrpIleSerP 181
580 CTGCTTCTGCTGCTGCTGCTA..... 602
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 roAlaPheCysGlnTrpPheValArgGluAlaValArgThrGlyArg 197
602 ..... 602
198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuPr 214
603 .....GGCCAT.....AATGGAATGT 618
214 ovalleuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTr 231
619 GGAGAAAGAACGTTCTTCTATGAGAACATCATGATGAGGAACAGAC 668
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
231 rPalGlySerThrArgSerThrHisThrGlySerSerCysIleGlyThrAsp 247
669 CTGAATAGCAACTTGTCTCCAAACACTGTGTGAGGAAGGTGCATCCAG 718
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 ProAsnArgAsnPhe... AspAlaGlyTrpPysGluIleGlyAlaSerAr 263
719 TTTCCTCATGCTCGGAACCTACTGTGAGCTTTATCCTGATGACAGACCA 768
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

263 gAsnProCysAspGlnThrTyrCysGlyProAlaAlaGlnSerGlnLysG 280
769 AAGTGAAAGCAGTGGCTAGTTCTTGAGAGAAAATATCCAGATTAATA 818
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
280 IuThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296
819 GCATACATCAGCATGATTCATCTACTCCAGCATATAGTGTTCATATTC 868
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
297 AlaTyrLeuThrIleHisSerLysSerGlnMetIleTyrProTyrSe 313
869 CTATACAGCAGTAAAGCAAGACCATGAGGAACTGCTCTAGTAGCCA 918
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 rTyAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaL 330
919 GTGAGCAGATCGTCTATGACAAACTAGTAATAAATATCCAGATATACA 968
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
330 yAlaIleThrValLysGluLeu... AlaSerLeuHisGlyThrLysTyrThr 345
969 CATGGCCATGGCTCAGAAACCTTATACCTGACCTCCGAGGTGGGAGCA 1018
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
346 TyrGlyProGlyAlaThrThrIleTyrProAlaAlaGlyLysSerAsp 362
1019 TTGGATCTATGATTTGGGACATCAATATGCTTACATC..... 1057
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
362 pTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPheGluLeuArg 379
1058 .....AAACCCACC 1066
379 sPThrGlyArgTyrGlyPheLeuLeuProGlnSerGlnIleArgAlaThr 395
1067 TGTAGAGAGCTTTTGCCGCTGTCTTAATAATAGCTTGCAATGATTAAG 1116
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
396 CysGlnGlnThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGln 412
1117 GAATGT 1123
412 ..... 414
seq_name: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1998.DAT:AAW74476
seq_documentation_block:
ID AAW74476 standard; protein; 415 AA.
XX
AC AAW74476;
XX
DE 20-NOV-1998 (first entry)
XX
DE Human pancreatic carboxypeptidase protein sequence.
XX
KW Human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KW prodruq therapy.
XX
OS Homo sapiens.
XX
PN WO9835988-A1.
XX
PD 20-AUG-1998.
XX
PE 10-FEB-1998; 98MO-GB00415.
XX
PR 29-OCT-1997; 97GB-0022727.
PR 14-FEB-1997; 97GB-0003104.
PR 18-OCT-1997; 97GB-0022003.
XX
PA (ZENE ) ZENEPHA (ZENE ) ZENEPHA LTD.
XX
PI Edge MD;
XX
DR MPI; 1998-467168/40.
DR N-PSDB; AAVal1795.
XX
PT New modified pro-domain of carboxy-peptidase B - enhances expression
of co-expressed proteins for production of recombinant

```

PT carboxy-peptidase or its fusions with antibodies, used, e.g. in
enzyme prodrug therapy

PS Example 1; Page 54-55; 83pp; English.

CC The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC modified pro-domain of CPB on a separate gene to enhance recombinant
CC expression. This process can be used to produce recombinant CPB in
CC eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
CC insulin production and protein sequencing, while its fusions with
CC antibodies are useful in antibody-directed enzyme prodrug therapy. The
CC modified pro-domain provide increased yields of recombinant CPB, possibly
CC by protecting the C-terminus against enzymatic degradation or by
CC increasing intracellular trafficking.

XX Sequence 415 AA;

Alignment_scores:
Quality: 651.50 Length: 419
Ratio: 2.404 Gaps: 8
Percent Similarity: 64.678 Percent Identity: 33.890

Alignment block:
US-09-980-881-1 x AAW74476 ..

Align seg 1/1 to: AAW74476 from: 1 to: 415

```

33 CTTCAGATCTTGTACCATTTGCTCTTGTGACAGCATGCTTCGC 82
||| ||||| :|||
2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisIsglyGly 18
83 G...TTCCAGATGCGCCAGTCTAGCTGCTTCTAGAACCTTAGGC 129
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
18 uHisPheGluGlyGlyValPheArgValAsnValGluAspGluAsn 35
130 AAGTCAAGTCTACAGAACTTACAAACATATGAGATGTCCTG 179
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
35 IsIleuAlaIleuArgGluLeuAlaSerThrThrGlnIleAspPheTrp 51
180 CAGCCGTAACAGCTGACCTTATGTGAAGAAACAAAGTCCATTTT 229
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
52 LysProAspSerValThrGlnIleLysProHisSerThrValAspPhe 68
230 TGTAAATGCATCTGATGTCGACAATGTGAAGCCCATTTAAATGAGCG 279
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
68 GValLysAlaGluAspThrValThrValGluAsnValLeuLysGluAsn 85
280 GAATTCATGACATGCTGTGCTGTCGACAGCGTGAAGATCTTATCAACG 329
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
85 IuLeuGlnIyrLysValIleuIleSerAsnLeuArgAsnValAlaGluAla 101
330 CAGATTTCCAGACACAGCTGACGCCCCGAGCCTCGCATGCTACTATGA 379
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
102 GlnPheAspSerArgVal.....ArgAlaThrGlyHisSerIyrGly 115
380 ACAGTATCATCTCACTAAATGAATCTATCTTGATAGAAATTTATACAG 429
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
115 uLysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaTrp 132
430 AGAGGATCCTGATGCTTACAAAATCCACATGTGATCCCTCATTTGAG 479
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
132 hGluAsnProAlaLeuIleSerArgSerValIleGlyThrTrpPheGln 148
480 AAGTACCCATCTATGTTTAAAGTTTCTGAAAAGAACAAACAGCCAA 529
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
149 GlyArgAlaIleTyrLeuLeuLysVal...GlyLysAlaGlyGlnAsnIyr 164
530 AATGCCATATGATGATGCTGGAATCCATGCGACAGAGATGATCTTC 579
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
164 sPheAlaIlePheMetAspCysGlyPheHisAlaArgIuTrpIleSerP 181
580 CTGCTTTCGCTTGCTGTGTCATTA..... 602

```

```

181 roAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197
602 ..... 602
198 GluIleGlnValThrGluLeuAspLysLeuAspPheTyrValLeuP 214
603 .....GGCCAT.....AATCGAATGT 618
214 ovalLeuAsnIleAspLyrTyrIleTyrThrTrpThrLysSerArgPhe 231
619 GAGAAAGAACCGTCTTCTATACGACAAATCATTCATCGGACAGAC 668
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
231 rPArgLysTrpArgSerThrHisThrGlySerCysIleGlyThrAsp 247
669 CTGATATGCAACTTGTCTCCAAACACTGGTGTGAGAGATGATCCAG 718
||| ||| ||| :||| :||| :||| :||| :||| :||| :|||
248 ProAsnArgAsnPhe...AspAlaGlyTrpCysGlnIleGlyAlaSer 263
719 TTCCATGCTCGAAGAACCTACTGTGACTTATCTGACTGACAAACAG 768
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
263 GAsnProCysAspGluThrTyrCysGlyProAlaAlaGluSerGluLys 280
769 AAGTGAAGCAGTGGCTAGTCTTGAGAAATATCAACGAGATTAA 818
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
280 IuThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerIleLys 296
819 GCATTCATCAGACATGATCATGATCCACATATAGTGTTCATATTC 868
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
297 AlATyrIleuThrIleHisSerTyrSerGlnMetCelleTyrTrpLys 313
313 rTyrAlaTyrLysLeuGlyGluAsnAsnAlaGluAlaLeuAlaL 330
919 GTAAGCAGTTCGCTATGTACAAAACACTAGTAAATATACAGTATACA 968
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
330 ysaLThrValLysGluLeu...AlaSerLeuHisGlyThrTyrTrp 345
969 CAGGCCATGCTCAGAAACCTTATACCTGCTGTCGAGAGTGGAGCA 1018
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
346 TyrGlyProGlyAlaThrThrIleTyrProAlaAlaGlyLysSerAsp 362
1019 TTGATCTATGATTTGGCATCAATATTCGTTACATC..... 1057
||| ||| ||| :||| :||| :||| :||| :||| :||| :|||
362 pTrpAlaTyrAspGlnGlyLeuArgTyrSerPheThrPheGluLeuArg 379
1058 .....AAACCACC 1066
379 sPThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395
1067 TGTAGAGAAAGCTTTGCGCTGTCTGTAATAATAGTTGGCATGTCAATAG 1116
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
396 CysGlnGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuG 412
1117 GAATGTT 1123
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
412 uHisLeu 414

```

seq_name: /SID1/gcgdelta/geneseq/geneseqp_emb1/AA1996.DAT:AAW06175

seq_documentation_block:

ID AAW06175 standard; Protein: 424 AA.

XX AAW06175;

XX AC

XX DT 17-FEB-1997 (first entry)

XX DE PROCPB with PeLB leader sequence.

XX KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPRT;

XX KW mustard-ribonucleotide; antibody directed enzyme prodrug therapy;

KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;

```

KM reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase
KW Fd; F(ab')2; PelB; leader; human carboxypeptidase B.
XX Synthetic.
OS
XX
XX Key location/Qualifiers
XX Key 1..22
XX Peptide /label= PelB_leader_sequence
XX FT 23..117
XX FT /label= pro_sequence
XX FT /label= 118..424
XX FT /label= mature_HCPB
XX
XX M09620011-A1.
XX
XX 04-JUL-1996.
XX
XX 21-DEC-1995; 95MO-GH02991.
XX
XX 16-AUG-1995; 95GB-0016810.
XX 23-DEC-1994; 94GB-0026192.
XX
XX (ZENE ) ZENECA LTD.
XX
XX Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DM;
XX Hennam JF, Henneguin LFA, Marsham PR, Rabin BR, Slater AM;
XX Tarragona-Fiol A, Taylorson CJ;
XX
XX WPI: 1996-321650/32.
XX DR
XX N-PSDB; AAT42506.
XX
XX Two component system for anti-tumour therapy - comprising targeting
XX PT moiety linked to mutated enzyme which can transform an
XX PT moiety:neoplastic produrg
XX
XX Reference Example 18; Page 140-142; 182pp; English.
XX
XX A two-component system for anti-tumour therapy comprises a targeting
XX CC moiety linked to a mutated enzyme which can transform an anti-neoplastic
XX CC produrg. The system is based on antibody directed enzyme produrg therapy
XX CC (ADEPT) using non-naturally occurring mutant forms of host enzymes,
XX CC pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
XX CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
XX CC be used. The present sequence is that of proHCPB with a PelB leader
XX CC sequence encoded by PIC11738.
XX
XX
XX Sequence 424 AA;
XX
XX
XX alignment_scores:
XX quality: 644.00 length: 401
XX Ratio: 2.458 gaps: 7
XX Percent Similarity: 65.337 Percent Identity: 33.915
XX
XX alignment_block:
XX US-09-980-881-1 x AAW06175 ..
XX
XX Align seg 1/1 to: AAW06175 from: 1 to: 424
XX
XX 84 TTCACGAGTGGCCAACTCTAGCTGCCTTCTTAGAACCTCTAGGCAACT 133
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 29 PhcgluglglglulysValPheargValAsnValGluaspGluAsnHisI 45
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 134 TCACATCTACAGAACTTACTACAAACATATAGAGATTGTCCTGGCAGC 183
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 45 eAsnIleIleArgIuIleuAlaSerThrThGlnIleAspHetrIlysp 62
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 184 CGGTAAACAGCTGACCTTATGTGGAAGAAAAAACAAGTCATTTTGTGTA 233
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 62 rGAspSerValThrGlnIleLysProHisSerThrValAspPheArgVal 78
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 234 AATGCTCTGATGTGACAAATGTGAAAGCCATTTTAAATGTGAACGGAA 283

```

79 LysAlGluAspThrValThrValGluAsnValLeuLysGlnAsnGluLeu 95
284 TCACATGCAGTCTCTTGTCGGCAGACGTGGAAAGATTATTCAACAGCAGA 333
 :::|||||:::|||||:::|||||:::|||||:::|||||:::
95 uGlntyrValLeuIleSerAsnLeuArgsnValValGluAlaGlnP 112
334 TTTCACAACGACACAGTCAGCCCCCGAAGCTCCGGCATCGACTATGACAG 383
 :::|||||:::|||||:::|||||:::|||||:::|||||:::
112 heAspSerArgVal.....ArgAlaThrGlyHisSerTyrGluLys 125
384 TATCACTCACATAAATGAATCATCTCTGGATAGATTAATTAACAGTAGAG 433
126 TyrAsnLysTrpGluThrIleGluAlaThrPrpThrGlnGlnValAlaThrG1 142
434 GCATCTCGATATGCTTTACAAAATAATCCACATTGGATCTCATATTGGAGAGT 483
 :::|||||:::|||||:::|||||:::|||||:::|||||:::
142 uAsnProAlaLeuIleSerArgSerValIleGlyThrThrProGluGluYLA 159
484 ACCCACTCTATGTTTTAAAGCTTTCTGGAAAAAGAACCAACGCCAAAT 533
 :::|||||:::|||||:::|||||:::|||||:::|||||:::
159 rGa1a1eTyLeuLeuLysVal...GlyLysAlaGlyGlnAsnLysPro 174
534 GCCATATGATGATGACTGGATCGGATCCGACGAGAGATGGATCTCCGCG 583
175 AlaIlePheMetAspCysGlyPheHisAlaArgLutIlePheSerProAl 191
584 TTTCTGCTTCTGCTGCATTA..... 602
208 IegInValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProVal 224
603GGCCAT.....AATCGAATGGCGAG 622
225 LeuAsnIleAspGlyTyrTylIeTyrThrTrpThrLysSerArgPheThrPar 241
623 AAGAACAACCGCTTCTTCTATGTCGCAACATCATGTCATCGGACACAGCTGA 672
241 gLyThrThrArgSerThrHisThrGlySerSerCysIleGlyThrAspProA 258
673 ATAGCAACTTTTCTCCAAACACAGCTGGTGTGAGGAAGGTGCATCCAGTTCC 722
 || |||||:::|||||:::|||||:::|||||:::|||||:::
258 snArGAsnPhe...AspIlaGlyTrpCysGluIleGlyAlaSerArgAsn 273
723 TCATGCTCGGAACCTACTGTGGACATTATCTCTGAGTGCAGAACGAAAT 772
274 ProCysAspGluThrThyCysGlyProAlaAlaGluSerGlnLysGluTh 290
773 GAAGCAGACTGGCTAGTATCTTGAGAGAAGAAATATCAACAGATTAAAGCAT 822
290 rLyAlaLeuAlaAspPheIleArgAsnLysLeuSerIleLysAlaAT 307
823 ACATCAGCATCATTCATATCTCCACAGATATAGTGTTCATATTCATAT 872
 || :||:::|||||:::|||||:::|||||:::|||||:::
307 yLeuThrIleHisSerTyrSerGlnMetIleTyrProTyrSerTyr 323
873 ACACGAGTAAAGAAAGACCATAGAGAACTGCTCTATGAGCCAGTGA 922
324 AlaTyrrLysGluGlyLysAsnAsnAlaGluLeuAsnAlaLeuAlaLysAl 340
923 AGCAGTTGCTGCTATATGCAAAAACATAGTAAATATCCAGGTATACAGT 972
340 aThrValLysGluLeu...AlaSerLeuHisGlyThrLysTyrThrTyG 356
973 GCGATGGCTCGAACAACCTTATACCTAGAGCTCCGAGAGTGGGAGCAATTGG 1023
356 LyrProGlyAlaThrThrLeyrProAlaAlaGlyLysAspAspTrp 372
1023 ATCTATGATTTGGGCGATCCAATATTCGTTTACATC..... 105
73 AlatyAspGlnGlyLysTyrSerPheThrPheGluLeuArgAspnh 389

```

1058 .....AAACCCACCTGTA 1070
      389 rGLyArGTyrgLYpHeLeuLeuProGLuSerGLnIleArGLaTAlThrCYG 406
      1071 GAGAGCTTTGCCGTCGTCCTAAATGCTTGCAATGTCATTAGGAA 1120
      406 lUGlUrThrPHeLeuAlaIleLysTYrValAlaSerTYrValLeuGLnHIS 422
      1121 GWT 1123
      423 Leu 423

seq_name: /SIS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.AAR75131
seq_documentation_block:
ID   AAR75131 standard; Protein; 404 AA.
XX
XX   AAR75131;
XX
XX   13-MAR-1996 (first entry)
XX
XX   Porcine Tyr-His-Met Procarboxypeptidase B.
XX
XX   Procarboxypeptidase B; carboxypeptidase B; Pichia; pCPB;
XX   human serum albumin; premating factor alpha; mating factor alpha;
XX   pOCBB.
XX
XX   Sus scrofa.
XX
XX   WO9514096-A1.
XX
XX   26-MAY-1995.
XX
XX   16-NOV-1994; 94WO-US13142.
XX
XX   16-NOV-1993; 93US-0153258.
XX
XX   (ELIL ) LILLY & CO ELI.
XX
XX   Fayerman JT, Greenen DP, Hersherberger CL, Larson JL;
XX   Sterner JL, Zhang H;
XX
XX   WPI: 1995-200386/26.
XX   N-PSDB; AAQ90600.
XX
XX   DNA encoding porcine carboxypeptidase B - used for transforming
XX   host cells, partic. Pichia species, for prodn. of the enzyme
XX
XX   Example 1; Page 16-17; 34pp; English.
XX
XX   The porcine carboxypeptidase B coding sequence can be place in a
XX   bacterial or pref. Pichia yeast expression vector. The expression
XX   vector further comprises the signal peptide of either human serum
XX   albumin (designated pSGD23 - NRRL B-21029); premating factor alpha
XX   (designated pF0489 - NRRL B-21028); mating factor alpha (designated
XX   pF0474 - NRRL B-21032) or the porcine pOCBB signal peptide;
XX   (designated pLGD27 - NRRL B-21027). The method can be used for
XX   producing large amounts of porcine carboxypeptidase B and when
XX   produced in Pichia yeast, the protein does not need solubilisation
XX   or folding. The produced enzyme is then used for pref. cleaving
XX   basic residues from the carboxy terminus of proteins.
XX
XX   Sequence 404 AA;

```

```

alignment_scores:
  Quality: 640.50      Length: 361
  Ratio: 2.680         Gaps: 6
  Percent Similarity: 66.205      Percent Identity: 37.673

```

```

alignment_block:
US-09-980-881-1 x AAR75131 ..

```

```

Align seg 1/1 to: AAR75131 from: 1 to: 404
84 TTCCACAGTGGCCAAATTCAGCTGCTCTCCATGACCTGATGAGCAAGT 133
10 PHeuGLIGlYGLySValPHeArGLaSnValGlaSPGLuSnAPIL 26
134 TCAAGTCTACAAATCTTACTACAACATATGAGATTGCTCTGCGACG 183
26 eSeRLeuHISGLuLeuAlaSerThrArGLnIleAspPHeTPrLysP 43
184 CGGTACACAGCTGATTTATGTCAGAAAAAACAGTCCATTTTTGTGA 233
43 rOAsPSeRValThrGLnIleLysProHISeRThrValAspPHeArGL 59
234 AATGCATCTGATGTCGACAATGGAAGCCATTAAATGTCGGCAGAT 283
60 LysAlaGLuAspPLeuAlaValAlaAspPHeLeuGLnHISnGLuLe 76
284 TCCATGCAAGTGTCTGTCGACAGCTGGAGATCTTATTCACACAGCA 333
76 uGLnTYrGLuValLeuAlaSnHISnLeuArGSeRValLeuGLuAlaGLn 93
334 TTTCACACGACAGTCAGCCCCGAGCCTCGCATCGTACTATGACAG 383
93 hEAsPSeRArGLVal.....ArgThrThrGLyHISeRTYrGLuLys 106
384 TATCACTACATGAATGAATCTATTCTGATAGATTAATTAACAGAG 433
107 TYrASnAsnTrpGLuThrIleGLuAlaTrpTrLysGLnValThSeRl 123
434 GCATCTGATATGCTTACAAAAATCCATGAGTCCATTTGAGAGAT 483
123 uAsnProAspPLeuIleSeRArGLAlaIleGLyThrThrPHeuGLyA 140
484 ACCCACTATGTTTAAAGTTTCTGGAAGAAGCAACAAAGCCAAAT 533
140 snAsnIleTYrLeuLeuLysVal...GLySPProGLyProAsnLysP 155
534 GCCATATGATGACTGTGGAATCCATGCCAGAAATGATGCTCTCC 583
156 AlaIlePHeMeLAspCysGLyPHeHISAlaArGLnTrpIleSeRHISAl 172
584 TTTGCTGTTGTGTCATA..... 602
172 aPHeCysGLnTrpPHeValArGLuAlaValLeuThrTYrGLyTYrGLuS 189
602 ..... 602
189 eRHISMeLThrGLuPHeLeuAsnLysLeuAspPHeTYrValLeuProVal 205
603 .....GCCAT.....AATCGAATGNGAG 622
206 LeuAsnIleAspGLyTYrIleTYrThrTrpTrLysAsnArGLMeLTrpAr 222
623 AAGAAGCGTTCTTCTATGCGAACATATGATCGAGACAGACCTGA 672
222 gLYSThrArGSeRThrAsnAlaGLyThrThrCysIleGLyThrAsPPro 239
673 ATAGCAACTTGTCTCCAAACACTGTGTGAGGAAGTGCATCCAGTTCC 722
239 snArGLsnPHe...AspAlaGLyTrpCysThrThrGLyAlaSeRThrAsP 254
723 TCATGCTCGAAACCTACTGTCATTTATCCGATCGACACCAAGAT 772
255 ProCysAspGLuThrTYrCysGLySeRAlaAlaGLuSerGLuLysGLuTh 271
773 GAAGCAGTGCCTACTTCTTGAGAGAAATATCAACAGATTAAGCAT 822
271 rLysAlaLeuAlaAspPHeIleArGLAsnHISnLeuSeRSeRlIleLysAl 288
823 ACATGACATGCATTCATCTCCAGCATATAGTGTTCATATTCCTAT 872

```

```

288 yIleuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSerTyr 304
873 ACACGAAGTAAAGCAAGACCATGAGAACTGCTCTAGTACCGACATGA 922
      ::::::::::::::::::::
305 AspTyrIleuProGluAsnAsnIleuLeuAsnIleuAlaTyrSAI 321
923 AGCAGTTCGCTATGTGACAAACTAGTAAATAACAGATACATG 972
      |||||:::
321 aAlaValIleuGluLeu...AlaThrLeuTyrGlyThrIleTyrThrTyrG 337
973 GCCATGCGCTGACAAACCTTATACCTACCTCCGAGGAGTGAGCATGG 1022
      |||:::
337 LyrProGlyAlaThrThrIleTyrProAlaIleGlyGlySerAspPirp 353
1023 ATCTAGATTGGCGCATCAATATTCGTTACA 1055
      |||||
354 AlaTyrAspGlnGlyIleIleTyrSerPheThr 364

seq_name: /stdd1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA66562
seq_documentation_block:
ID   AA66562 standard; Protein; 324 AA.
XX
AC   AA66562;
XX
DT   22-OCT-2001 (first entry)
XX
DE   Bothops jararaca carboxypeptidase homologue protein.
XX
KW   secreted metallocarboxypeptidase; immunomodulatory;
KW   antiparkinsonian; nootropic; anticonvulsant; neuroprotective;
KW   osteoparalic; antiarthritic; cerebroprotective; vasotropic; vulnery;
KW   antibacterial; antiviral; antifungal; antiinflammatory; antidiabetic;
KW   antineumatic; cyostatic; antiallergic; haemostatic; infection;
XX   cancer.
XX
OS   Bothops jararaca.
XX
PN   WO200157265-A1.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03783.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   29-SEP-2000; 2000US-0676135.
XX
PA   (HYPE-) HYSEQ INC.
XX
PI   Boyle BJ, Mize NK, Arterburn MC, Yeung G, Tang YT, Liu C;
PI   Drmanac RT, Wang M, Chen L, Yang Y;
XX
DR   WPI; 2001-465705/50.
XX
PT   Novel metallocarboxy peptidase-like polypeptide is useful in treating
PT   neurodegenerative, immune, etc. disorders
XX
PS   Example 3; Fig 1; 146pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino
CC   acid sequence of 374, 41, 14, 15, 41, 22, 15, 27, 14, 9, 18, 20, 15,
CC   354, 165, 374, 354, or 41 amino acids, fully defined in the
CC   specification, or its translated protein coding portion. The protein
CC   coding portion, extracellular portion or active domain. The polypeptide
CC   is a human secreted metallocarboxypeptidase-like polypeptide. It
CC   is useful for treating mammalian diseases including:
CC   neurodegenerative diseases such as Parkinson's or Alzheimer's diseases,
CC   osteoporosis, osteoarthritis, disorders linked to reduced tissue
CC   growth, Huntington's disease, amyotrophic lateral sclerosis, stroke,
CC   reperfusion, wound healing, infectious diseases, immune and autoimmune
CC   disorders such as severe combined immunodeficiency, rheumatoid
CC   arthritis, Guillain-Barre syndrome, insulin dependent diabetes

```

```

CC   mellitus, graft-versus-host disease and allergy, cancer, and
CC   haemophilia. The polypeptide is also useful in medical imaging, as an
CC   immunogen to generate antibodies, as a nutritional source or
CC   supplement, and as a contraceptive. The present sequence is
CC   a carboxypeptidase homologue protein of Bothops jararaca.
XX
SQ   Sequence 324 AA:

alignment_scores:
      Quality: 636.00      Length: 329
      Ratio: 2.917      Gaps: 6
Percent Similarity: 66.261      Percent Identity: 40.122

alignment_block:
US-09-980-881-1 x AA66562 ..

Align seg 1/1 to: AA66562 from: 1 to: 324

267 TTAAATGTGACGGAAATTCATGACAGTGTCTTGTGCGACAGCTGGAAGA 316
      |||||
2   LeuGlnGlnSerGlyLeuAsnTyrGluIleuIleAsnLeuGlnAl 18
317 TCTTATTCACAGCAGATTTCACAGACAGTACGCCCGAGCTCCG 366
      ::::::::::::::::::::
18 aValLeuAspArgGlnLeuAspAsnHisAla.....ArgThrAlaG 32
367 CATGCTACTAGACAGATTCACATCAGTAAATGAAATCTATCTGGATA 416
      ::|
32 LyrTyrAsnTyrGluTyrTyrAsnSerTrpGluTyrIleAspAlaThrPir 48
417 GAATTTATTAAGTACAGAGCATCCTGATATGCTTACAAATTCACATGG 466
      |||||
49 AlaAspIleAlaAsnGluAsnProSerIleuValSerArgLeuGlnIleG 65
467 ATCCGATTTGAGAAAGTACCCACTATGTTTAAAGTTTCTGGAAGA 516
      ::::::::::::::::::::
65 YThrThrPheGlnGluArgProMetProLeuLeuVal...GlyLysP 81
517 AACAAACAGCCAAATATGCAATGATGATGATGATGATGATGATGATG 566
      ::|
81 roGlyValAsnIleValAsnIlePheIleAspCysGlyPheHisAlaArg 97
567 GAATGATCTCTCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCT 602
      |||||
98 GluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValAr 114
602 .....
114 gThrTyrGlyLysGluThrIleMetThrGlnLeuLeuAsnIleuAspP 131
603 .....GSC 605
131 heTyrIleLeuProValLeuAsnIleAspGlyTyrValTyrSerTyrLys 147
606 CATAAATGCAATGTGAGAAAGAACCGTCTTTCTTATGGAACATCATG 655
      ::::::::::::::::::::
148 GlnSerArgMetTrpArgLysThrArgSerValAsnAlaGlySerThrCy 164
656 CATGGAACAGACCTGAATAGCACTTGTCTCCAAACACAGCTGTGAGG 705
      |||||
164 sIleGlyThrAspProAsnArgAsnPhe...AspAlaIleTrpCysSerV 180
706 AAGGCGATTCAGCTTCATGCTGCGGAACCTACTGATGATTAATCT 755
      |||||
180 aGlyAlaSerArgAsnProCysSerGluThrTyrCysGlySerTyrPro 196
756 GAGTCAGAACCAAGAGTGAAGGCACTGCTTATCTTGAGAAAGAAATAT 805
      |||||
197 GlnSerGluTyrGluThrTyrAlaLeuAlaAspPheIleArgArgAsnAr 213
806 CAACAGATTAAGCATATCATCATGATCATCATCATCATCATCATCAT 855
      ::::::::::::::::::::

```

213 gSerIleIleGlnAlaTyrLeuThrIleHisSerTyrSerGlnMetLeuL 230
 856 TGTTTCATATTCCTATACAGAGTAAAGCAAGCATGAGAGACTG 905
 230 eutryProtyrSerTyrThrTyrAspLeuThrSerAsnAnuLysLeu 246
 906 TCCTAGTACGAGTACAGAGCTTGGCTATTGACAAACTGATAAAA 955
 247 AsnSerIleAlaLysGlnAlaIleArgGlnLeu...LysValLeuPheG 262
 956 TACAGGTATACATAGCCATGCTCAGAAACCTTATACCTGCTCTG 1005
 262 ymHrGluTyrThrTyrGlyProGlyAlaAlaThrIleTyrProAlaIaG 279
 1006 GAGGTGGGAGCATGATGATCTATGATTTGGGATCAAAATATGCTTACA 1055
 279 LysIleSerAspAspTrpAlaTyrAspGlnIleLysTyrAlaPheThr 295
 1056 TC..... 1057
 296 PheGlnLeuArgAspLysGlyArgTyrGlyPheAlaLeuProGlnSerG 312
 1058AAACCCACTGTAGAGACCTTTGGCGCTGC 1090
 312 nileLysProThrCysGlnIleuThrMetIleAlaVal 324
 seq_name: /SID51/cgdata/geneseq/genesep-emb1/AA2001.DAT:AAE00502
 seq_documentation_block:
 ID AAE00502 standard; Protein; 437 AA.
 AC AAE00502;
 DT 19-JUN-2001 (first entry)
 XX Human carboxypeptidase-related protein #5.
 DE Human carboxypeptidase-related protein #5.
 KW Human; carboxypeptidase-related protein; protease; gene therapy;
 KM drug screening; mental; biological disorder; medical disorder.
 OS Homo sapiens.
 PN WO200123588-A1.
 PD 05-APR-2001.
 PF 29-SEP-2000; 2000WO-US26876.
 PR 29-SEP-1999; 99US-0156685.
 XX (LEXI-) LEXICON GENETICS INC.
 PA Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 PI WPI: 2001-266171/27.
 DR N-PSDB: AAD03837.
 XX New isolated human carboxypeptidase polynucleotide useful for
 PT generating antibodies, as reagents in diagnostic assays and for
 screening for compounds useful for treating mental, biological or
 PT medical diseases -
 XX
 PS Claim 2; Page 30-31; 36pp: English.
 CC The present sequence is a novel human protein (NHP),
 CC carboxypeptidase-related protein. The carboxypeptidase-related
 CC proteins share structural similarity with animal proteases, and
 CC especially carboxypeptidase B or carboxypeptidase A. The
 CC carboxypeptidase-related protein cDNA is useful for the detection of
 CC mutant human carboxypeptidases, or inappropriately expressed human
 CC carboxypeptidases for the diagnosis of disease, and also as a
 CC therapeutic. It is also useful in drug screening, for generation of
 CC antibodies, as reagents in diagnostic assays, for the identification of

CC other cellular gene products related to human carboxypeptidases, and as
 CC reagents in assays for screening for compounds that are useful for
 CC treating mental, biological or medical disorders and diseases. Nucleotide
 CC constructs encoding functional NHPs can also be used in gene therapy.
 XX

SD Sequence 437 AA:

alignment_scores:
 Quality: 635.50 Length: 437
 Ratio: 2.380 Gaps: 12
 Percent Similarity: 61.098 Percent Identity: 33.638

alignment_block:
 US-09-980-881-1 x AAE00502 ..

Align seg 1/1 to: AAE00502 from: 1 to: 437

9 GCCTGTGGGATGAAGCTTTGC.....AGCCTTGCAAGCTCTGTACCCAT 52
 12 AlAlaPheLeuProLeuGlyStrLeuPheLeuLysIleLeuGlnPro.. 27
 53 TGTTCTTCTGTGAGCAGCATGCTTC.....GCCTTCAGAGTGGCC 96
 28GlyHisSerHisLeuTyrAsnAsnArgTyrAlaGlyAspL 41
 97 AAGTTCAGCTGCTCTCTAGAACCTCTAGCAAGTCAAGTTCTACAG 146
 41 yValIleArgPheIleProLysThrGlnGlnAlaTyrAlaLeuLys 57
 147 AATCTTACTACAAACATATGATGCTCTCTGCGACGCCGTAACAGCTGA 196
 58 LysIleSerTyrGlnLeuLysValAspLeuTrpGlnProSerSerIleSe 74
 197 CCTTATGTGAG.....AAAAACAAGTCATTTTGTGTAAGCAT 240
 74 TrYValSerGlnGlyThrValThrAspValHisIleProGlnAsnLys 91
 241 CTGATGTCGCAAAATGTGMAAGCCCATTTAATGTGAGCGGAATTCATGC 290
 91 eArgAla.....LeuLeuAlaPheLeuGlnGlnAlaAsnIleGlnTyr 105
 291 AGTGTCTTGTGCGACAGCTGGAAGATCTTATTCACACAGCATTTCCAA 340
 106 LysValLeuIleGlnLysLeuGlnLysThrLeuGlnLysGlySerSerIle 122
 341 CGACACAGTCAGCCCGGAGCTCCGATGC.....TACTATGACAGT 384
 122 uHisThrGlnArgAsnArgSerLeuSerGlyTyrAsnTyrGlnValT 139
 385 ATCACTCACTAAATGAATCTATCTTGATAGAAATTAATCACTGAGAG 434
 139 yHisSerLeuGlnGlnIleGlnAsnTrpMetHisIleAsnAnuLysThr 155
 435 CATCTGATATGCTTACAAAAATCCACATTTGATCCTCATTTGAGAAGTA 484
 156 HisSerGlyLeuIleHisMetPheSerIleGlyArgSerTyrGlnLysArg 172
 485 CCACCTGATGTTTAAAGTTCTGTGAAAGAAACAACGCAAAAATG 534
 172 gSerLeuPheIleLeuLysLeu...GlyArgSerArgLeuLysArg 188
 535 CCATATGATGATGACTGTGAAATCCATGCCAGAAATGATCTCTCTGCT 584
 188 lAlaTrpIleAspCysGlyIleHisAlaArgLysTrpIleGlyProAla 204
 585 TTCTGCTTGGTTCATA..... 602
 205 PheCysGlnTrpPheValLysGlnAlaLeuLeuThrTyrLysSerAsp 221
 602 602
 221 oAlaMetArgLysMetLeuAsnHisLeuTyrPheTyrIleMetProValP 238

```

603 .....GGCCATTAATCGAATGTGGAGA 623
238 heasvalaspGLyTyrHisPheSerTrpThrAsnAspArgPheTrpArg 254
624 AAGAACGGTTCTTCTATGCGAACATCATCGGACGACAGACCTGAA 673
255 LysThrArgSerArgAsnSerArgPheArgCysArgGlyValAspAlaAs 271
674 TAGCACTTGTCTCCAAACACGTGTGAGAGAGTGATCCAGTTCT 723
271 nArgAsnTrpArgVallys...TrpCysAspGluGlyAlaSerMetHisP 287
724 CATCTGGGAAACCTACTGTGAGATTATCTGTAGTCAGAACCGAAGTG 773
287 roCysAspAspTrpTyrCysGlyProPheProGluSerGluProGluVal 303
774 AAGCAGGAGGTACTTCTTGAGAGAAATATCAACGAGTAAAGCTA 823
304 LysAlaValAlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTy 320
824 CATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATATTCCTATA 873
320 rLeuSerPheHisAlaTyrAlaGluMetLeuLeuTyrProTyrSerTyrL 337
874 CACGAACTAAAGCAAGACCATGAGACTGTCTTAGTCCAGTGA 923
337 ySTyrAlaThrIleProAsnPheArgCysValGluSerAlaAlaTyrLys 353
924 GCAGTGTGCTATTGACAAACATAGTAAATACGAGGTATACACATNG 973
354 AlaAlaValAlaLeu...GlnSerValTyrGlyValAlaGlyTyrGly 369
974 CCATGAGCTCAGAAACCTATCTAGCTAGCTCGAGAGTGAGGAGCATGGA 1023
369 yProAlaSerThrThrLeuTyrValSerSerGlySerSerMetHisPrrpa 386
1024 TCTATGATTGGCATCAATATTCGTTAC..... 1054
386 lATyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThr 402
1055 .....ATCAACCCACCTGAG 1071
403 GlyTyrPheGlyPheLeuLeuProGluMetLeuIleLysProThrCysTh 419
1072 AGAAGCTTTTGGCCGCTGTCTTAATAATAGCTTGGCATGTCAATTAGGAATG 1121
419 rGluThrMetLeuAlaValLysAsnIleThrMetHisLeuLeuLysLys. 435
1122 TTTAATGCCCC 1132
436 .....CysPro 437
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW41414
seq_documentation_block:
ID AAW41414 standard; Protein; 716 AA.
XX
XX AAW41414;
AC
XX 02-JUN-1998 (first entry)
XX
XX PreproHCPB-linker-humanised 806.077 VH.
XX
XX Anti-CEA antibody; carcinoembryonic antigen; 806.077 Ab; cancer therapy;
XX cancer diagnosis; complementarity determining region.
XX
XX Synthetic.
XX
XX WO9742329-A1.
XX
XX 13-NOV-1997.
XX

```

```

PF 29-APR-1997; 97MO-GB01165.
XX
XX 14-FEB-1997; 97GB-0003103.
PR 04-MAY-1996; 96GB-0009405.
XX
XX (ZENEC ) ZENECA LTD.
XX
XX Copley CG, Edge MD, Emery SC;
XX
XX WPI; 1997-558987/51.
DR N-PSDB; AAV17331.
XX
PT Anti-carcinoembryonic antigen antibody 806.077 Ab - used for
PT diagnosis and therapy of cancer
XX
PS Example 75; Page 184-187; 208pp; English.
XX
CC This sequence is preproHCPB-linker-humanised 807.077 VH, which is
CC an example of the antibody of the invention. The antibody is an anti-CEA
CC (carcinoembryonic antigen) antibody (preferably 806.077 Ab). Host cells
CC or transgenic organisms transformed with DNA encoding the antibody, are
CC used to make the antibody or conjugate. The conjugate is used in a
CC medicament suitable for intravenous administration. The conjugate can be
CC used for cancer therapy, selectively killing tumour cells. The antibody
CC can be used for in vivo or in vitro diagnosis of cancer.
XX
SQ Sequence 716 AA;

```

```

alignment_scores:
    Quality: 635.50      Length: 422
    Ratio: 2.345         Gaps: 8
Percent Similarity: 64.218      Percent Identity: 32.938

```

alignment_block:

US-09-980-881-1 x AAW41414 ..

Align seg 1/1 to: AAW41414 from: 1 to: 716

```

24 CTTTGCAGCCTTGGCAGCTCTGTACCATGTGTCTCTGTGAGACGA 73
1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisH 17
74 TGTCTTCGCG...TTCAGAGTGGCCAAATCTTACGCTCTTCTTAGAA 120
17 sGlyGlyGluHisPheGlyGlyGlyValAlaPheArgValAsnValGlu 34
121 CCTTCAGCAGTTCAGTTCATAGCATCTTACACAACTATGAGATT 170
34 sPGLuAsnHisIleAsnIleIleArgGluLeuAlaSerThrThrIle 50
171 GTTCTCTGCGACCCGCTAAGCTGACCTTATTTGAGAGAAAAACAGT 220
51 AsprhetrPlysProAspSerValThrGlnIleLysProHisSerThrVa 67
221 CCATTTTGTGTAATGCATCTGATGCGACAAATGGAAGCCATTGAA 270
67 lAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeuL 84
271 ATGTGAGCGGAATTCATGCAAGTGTCTGTGGCGAGAGGTGGAAGATCT 320
84 ySGLuAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100
321 APTTCAACAGCAGATTTCCACAGACAGCTCAGCCCGCAGCCTCGCATC 370
101 ValGluAlaGlnPheAspSerArgVal.....ArgAlaThrGlyH 114
371 GTACTATGACAGTATCACTCACTAATGAATCTATCTTGATAGAT 420
114 sSerTyrGluLysTyrAsnLysTrpGluThrIleGluAlaTrpThrGing 131
421 TTATACAGAGAGCATCTGATATGCTTACAAAATGCACATGAGATCC 470

```

```
131 lnValAlaThrGluAsnProAlaLeuIleSerArgSerValIleGlyThr 147
471 TCATTGAGAGACTACCCACTATATGTTTAAAGCTTTCGAAAGACACA 520
      ::::::::::::::::::::: |||||
148 ThrPheGluGlyArgAlaIleTyrIleuLeuLysVal...GlyLysAlaG1 163
521 AACAGCCAAAATGCGCATATGATTGATTCGTGAAATCATGCCAGAGAA 570
      ||| ||||| ::::::::::::::::::::: |||||
163 yGlnAsnLysProAlaIleIlePheMetLysPcysGlyPheHisAlaArgGlu 180
571 GCATCTCTCCCTGCTTCTGCTGTGGTCATA..... 602
      ||||| ::::::::::::::::::::: |||||
180 rPleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196
602 ..... 602
197 TyrGlyArgGluIleGlnValThrGluLeuAspLysLeuAspPheTyr 213
603 .....GCCCAT.....A 609
      ||:::
213 lValLeuProValLeuAsnIleAspGlyTyrIleTyrThrTrpLys 230
610 ATCGAATGTGGAGAAAGAACCCCTTTCATGCCAACAATCATTCGATC 659
      ::::::::::::::::::::: ::::::::::::::::::::: |||||
230 eArGpPheTrpArgLysThrArgSerThrHisThrLysSerSerCysIle 246
660 GGAACGACCTGAATAGCAACTTGTCTCCAAACACTGTGTGAGAGAG 709
      ||||| ||| ||||| ::::::::::::::::::::: ||
247 GlyThrAspProAsnArgAsnPhe...AspAlaGlyTrpCysGluIleG1 262
710 TGCATCCAGTTCTCATGCTCGAAMACTAGTGGACTTATCTGAGT 759
      ||||| ::| ::::::::::::::::::::: |||||
262 yAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 279
760 CAGAACCAGAGTGAAGCGCAGTGGCTAGTTCTTGAGAGAAATATCAAC 809
      ||||| ||||| ::::::::::::::::::::: ||:::::::::::
279 eRlIuLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295
810 CAGATTAAACATATACATGATCATCTTCACTCCAGCATATAGTGT 859
      ::::::::::::::::::::: |||||
296 SerIleLysAlaTyrIleuThrIleHisSerTyrSerGlnMetIleTyr 312
860 TCATATTCCTATACAGAAATAAAGCAAGACATGAGAACTGCTC 909
      ::::::::::::::::::::: ::::::::::::::::::::: |||||
312 rProTyrSerTyrAlaTyrLysLeuGlyLysAsnAsnAlaGluLeuAsnA 329
910 TAGAGCCAGTGAAGCAGTTCGTCTATGACAAAACACTAGTAAATATAC 959
      ::::::::::::::::::::: ::| ::| ::|
329 lAlaValAlaLysAlaThrValLysGluLeu...AlaSerLeuHisGlyThr 344
960 AGTATATACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAG 1009
      ::::::::::::::::::::: ||::| ::| ::|
345 LysTyrThrTyrGlyProGlyAlaThrThrIleTyrProSerAlaGlyThr 361
1010 TGGGGAGCATTTGATCTATGATTGGGCAATATTCGTTTACATC.. 1057
      ::| ||||| ||||| ||||| ::::::::::::::::::::: ||
361 rSerLysAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPheC 378
1057 ..... 1057
378 lIuLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394
1058 AAACCCACTGTAGAGAACTTTGCCGCTGCTCTTAAATAGCTTGCA 1107
      ::| ||||| ::| ||||| ||||| ::|
395 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTy 411
1108 TGTGATTAGGAATGT 1123
411 rValLeuGluHisLeu 416
```


OM of: US-09-980-881-1 to: Issued_Patents_AA:* out-format : pfs

Date: Sep 18, 2002 4:39 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+g2p.model -DEV=x1h
-O=/cgn2_1/USPRO_POOL/US09980881/runat_16092002_140039_8839/app-query.fasta_1.1655
-DB-issued_patents_AA -QEMT-fastan -SUFFIX=rai -CAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPTOL=0.000 -LOOEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORF=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881 -CGN1_1 14 -NCPU=6 -ICPU=3 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLIFY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-1
Query length: 1573
Database: Issued_Patents_AA:*
Database sequences: 231628
Database length: 24425594
Search time (sec): 34.320000

```
score_list:
Sequence      Strd Orig      ZScore      EScore Len 1 Documentation
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-3 + 1885.00 3732.52 4.8e-201 423 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-3 + 1885.00 3732.52 4.8e-201 423 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-4 + 1885.00 3732.52 4.8e-201 423 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-869-057-2 + 1885.00 3732.52 4.8e-201 423 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-7 + 655.00 1276.60 3.0e-64 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-7 + 655.00 1276.60 3.0e-64 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-7 + 655.00 1276.60 3.0e-64 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-869-057-2 + 655.00 1276.60 3.0e-64 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-6 + 647.50 1261.62 2.1e-63 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-6 + 647.50 1261.62 2.1e-63 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-6 + 647.50 1261.62 2.1e-63 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-696-139-2 + 640.50 1248.01 1.2e-62 404 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-09-171-945-125 + 634.50 1229.42 7.5e-62 716 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-782-760-6 + 607.50 1185.29 5.0e-59 307 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:PCR-US96-00995-6 + 607.50 1185.29 5.0e-59 307 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-4 + 607.00 1181.35 6.5e-59 396 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-4 + 607.00 1181.35 6.5e-59 396 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-4 + 607.00 1181.35 6.5e-59 396 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-696-139-4 + 576.00 1122.43 1.6e-55 306 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-09-171-945-113 + 571.50 1105.41 7.1e-55 613 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-8 + 553.00 1072.92 6.7e-53 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-8 + 553.00 1072.92 6.7e-53 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-8 + 553.00 1072.92 6.7e-53 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-5 + 529.00 1024.95 3.2e-50 419 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-5 + 529.00 1024.95 3.2e-50 419 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-5 + 529.00 1024.95 3.2e-50 419 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-782-760-4 + 513.00 993.05 1.9e-48 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-09-395-936-4 + 513.00 993.05 1.9e-48 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-640-906-18 + 512.00 991.05 2.4e-48 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-09-395-936-18 + 512.00 991.05 2.4e-48 417 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-640-906-2 + 487.50 942.08 1.3e-45 419 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-09-395-936-2 + 487.50 942.08 1.3e-45 419 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-640-906-17 + 484.00 935.09 3.2e-45 419 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-09-395-936-17 + 484.00 935.09 3.2e-45 419 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-1 + 167.00 330.14 1.8e-10 37 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-1 + 167.00 330.14 1.8e-10 37 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-1 + 167.00 330.14 1.8e-10 37 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-484-105-16 + 118.50 199.22 0.0002 706 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-484-105-16 + 118.50 199.22 0.0002 706 1
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-782-760-3 + 100.50 186.45 0.0070 95 1
```

```
seq_name: /cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-3
seq_documentation_block:
: Sequence 3, Application US/07649591B
: Patent No. 5206161
: GENERAL INFORMATION:
: APPLICANT: Dennis Drayna and Daniel Eaton
: TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/649,591B
: FILING DATE: 19910201
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 689
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/266-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 423 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: US-07-649-591B-3

alignment_scores:
Quality: 1885.00 Length: 423
Ratio: 5.122 Gaps: 2
Percent Similarity: 86.998 Percent Identity: 86.525

alignment_block:
US-09-980-881-1 x US-07-649-591B-3 ..
Align seg 1/1 to: US-07-649-591B-3 from: 1 to: 423

18 ATGAAGCTTTCGACGCTTCAGCTTCCTGACCTGCTTCCTTGTGCA 67
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-3
1 Mettlyslaucysserleualavalalleuvalproilleualleuphecysgl 17
68 GGAGATGCTTCGCTTCAGAGGCGCAAGTTCAGCTGCTTCCTGCA 117
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-3
17 ucinihsvalpnealapegnlsersglvalleualalaaleuproA 34
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-3
118 GAACCTTCAGGCAAGTTCAGATCTTACAGAACTTACTACAACTATGAG 167
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-3
34 rgtfhsrarglnvalglnvalleuglnasrleuthrthrrhyrfglu 50
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-3
168 ATTGTTCTGTGCGACCGGTAACAGCTGACTTATTGAGAGAAAAACA 217
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-3
```

```

|||||
51 lIeValleuPrGlnProValThrAlaAspLeuIleValLysLysLysG 67
218 AGTSCATTTTGTGAAATGCATGTGATGCGACAATGTGAAGCCATT 267
|||||
67 nValnIsPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
268 TAAATGTGAGCGGAATTCATGCAGTGTCTTGTCGACAGCTGGAAGAT 317
|||||
84 euAnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
318 CTATATCAACAGCATTTCCACAGACAGTCAGCCCGGACCCCTCCG 367
|||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaSerAl 117
368 ATGCTACTATGACAGTATGCTACTGTAATGAAATCTATCTTGATAG 417
|||||
117 aserTyTYTgLuGlnTyThrIleSerLeuAsnGluIleTySerTrpIleG 134
418 AATTATTAACAGAGGAGCATCTGATATGCTTAACAATAATCCACATTGA 467
|||||
134 lPheIleThrGluArgHisProAspPheLeuThrLysIleHisIleGly 150
468 TCCTCATTTGAGAAAGTACCCACTATGCTTTTAAAGTTCTGGAAGA 517
|||||
151 SerSerPheGluLysTyTrpLeuTyValLeuLysValSerGlyLysG 167
518 ACAAAAGCCAAATAATGCATATGATGATGCTGGAATCCATGCGAGAG 567
|||||
167 uGlnThrAlaLysAsnAlaIleThrIleAspCysGlyIleHisAlaArg 184
568 AATGATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608
|||||
184 lTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
608 ..... 608
201 pheTyGlyIleIleGlyGlnTyThrAsnLeuLeuArgLeuValAsp 217
608 ..... 608
217 eTyValMetProValValAsnValAspGlyTyAspTySerTrpLysL 234
609 . AATGCAATGTGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTC 656
|||||
234 ysAsnArgMetIleTrpArgLysAsnArgSerPheTyThrAlaAsnHisCys 250
657 ATCGAAGACAGCTGAATAGCACTTGTCTCCAACACACTGAGTGTGAGA 706
|||||
251 lIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisIleTrpCysGlu 267
707 AGGTGATCCAGTTCTCTGCTGCGAACAACCTACTGTGCACTTATCTG 756
|||||
267 uGlyAlaSerSerSerSerCysSerGlnTyTrpCysGlyLeuTyProG 284
757 AGTCACAACAGAGAGCGAGCGAGTGTGCTGCTGCTGCTGCTGCTGCTG 806
|||||
284 luserGlnProGlnValLysAlaValAlaSerPheLeuArgAsnIle 300
807 AACACAGATTAAGCATACATGACATGATGCTTACTCCAGCATATAGT 856
|||||
301 AsnGlnIleLysAlaTyThrIleSerMetHisSerTySerGlnHisIleVa 317
857 GTTTCATATTCCTATACAGAAATAAAAGCAAGACATGAGAACTGT 906
|||||
317 lPheProTySerTyThrArgSerLysSerLysAspHisGluGlnLeuS 334
907 CTCTAGTACCGAGTGAAGCACTTCGCTATTGACAAAAGTAAATAAT 956
|||||
334 eTyLeuValAlaSerGlnAlaValArgAlaIleGlyLysThrSerLysAsn 350
957 ACCAGGTATACATGCGCATGCGTCAGAAACCTTACTACTCTGCTG 1006
|||||

```

```

351 ThrArgTyThrHisGlyHisGlySerGluThrLeuTyLeuAlaProG 367
1007 AGGTGGGAGCATGTGATGATGATGATTTGGGCATCAATATTCGT ..... 1051
|||||
367 yGlyGlyAspAspTrpIleTyThrAspLeuGlyIleLysTySerPheThr 384
1052 .....TAC 1054
384 leGluLeuArgAspThrGlyTyThrGlyPheLeuLeuProGluArgTy 400
1055 ATCAAAACCCAGCTGTAGAGACGTTTGGCGTGTCTCTAAATAGCTG 1104
|||||
401 lIeLysProThrCysArgGluAlaPheAlaValAlaValSerLysIleAla 417
1105 GCATGTCATAGCAATGTT 1123
|||||
417 pHisValIleArgAsnVal 423

```

seq_name: /cgn2_6/plodata/1/laa/5A.COMB.pep:US-08-277-540-3

seq_documentation_block:

; Sequence 3, Application us/08277540

; Patent No. 5474901

; GENERAL INFORMATION:

; APPLICANT: Drayna, Dennis T., Eaton, Dan L.

; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: patln (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/277,540

; FILING DATE: 19-JUL-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/167727

; FILING DATE: 15-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/959944

; FILING DATE: 14-OCT-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/649591

; FILING DATE: 01-FEB-91

; ATTORNEY/AGENT INFORMATION:

; NAME: Hasak, Janet E.

; REGISTRATION NUMBER: 28,616

; REFERENCE/DOCKET NUMBER: 689D1C1D1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/225-1896

; TELEFAX: 415/952-9881

; TELEX: 910/371-7168

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 423 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; US-08-277-540-3

alignment_scores:

Quality: 1885.00

Ratio: 5.122

Percent Similarity: 86.998

Length: 423

Gaps: 2

Percent Identity: 86.525

alignment_block:
ms-09-990-991-1

US-09-980-881-1 x US-08-277-540-3

Align seg 1/1 to: US-08-277-540-3 from: 1 to: 423

18	ATGAACTTTGAGAGCTTGCACCTGCTGTAACCAATGTCCTCTGTGA	67
1	MetLysLeuGlySerLeuAlaValLeuValProIleValLeuPheCysG	17
68	GAGCACTGCTTCCGCTTCCAGAGTGGCCAAATTCTACTGCTCTTCTA	117
17	UGlnIHisValPheAlaPheGlnSerGlyGlnValLeuAlaIleuProA	34
118	GAACCTCTAGGCAAGTTCAAGTTCTACAGAACTTTACTACACATATGAC	167
34	rgthSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGlu	50
168	ATTGTTCTCGGAGCGGCTTAACCTGCAACCTTATGTCGAAGAAACA	217
51	IleValLeuThrPheGlnProValThrAlaSerLeuIleValLysLysG	67
218	AGTCCATTTTGTGTAATGCACTGATGTGCGACAAATGTAAGCCCAAT	267
67	ValIHisPhePheValAsnAlaSerAspValAlaSerValLysAlaHisL	84
268	TAATGTGACCGCAATTCATGCACTGCTGTGTCGGACGTGGAAAT	317
84	eunAlaValSerGlyIleProCysSerValLeuLeuAlaAspValGlnAsp	100
318	CTTATTTCACAGACAGATTTCCCAACACACAGTCAACGCGCGGCTCCG	367
101	LeuIleGlnGlnGlnIleSerAsnSphThrValSerProAlaGlnSerAl	117
368	ATGCACTATGAAAGTACTACTACTAAATGAATATATTCTTGATAG	417
117	asertTyrGlnUGlnTyrHisSerLeuAsnGlnIleTyrSerTrpIleG	134
418	AATTTTTAATCTGAGGAGCACTCGATATGCTTAAACAAAATCCACATTTGA	467
134	IubelleIethrGlnArgHisPheProAspMetLeuThrLysIleHisIleGly	156
468	TCTCATTTTGAGAGTACCCACTCATGTGTTTAAAGTTTCTGGAAAGA	517
151	SerSerPheGlnLysTyrProLeuTyrAlleuLysValSerGlyLysG	167
518	ACAAACAGCCAAAATGCCATATGATTTGATGTGGAAATCCATGCCAGAG	567
167	UGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg	184
568	AATGCAATCTCCCTGTTTCTGCTGTGGTTCATATGAGCAAT	608
184	IutTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln	200
608	608
201	PheTyrGlyIleIleGlyGlnTyrThrAsnLeuIleuArgLeuValAspRh	217
608	608
217	eTyrValMetProValValAsnValAspGlyTyrAspTyrSerTrpLysL	234
609	..AATGCAATGTGTGAAAGAAGACCGTCTCTTCTATGCAACAATCATTTGC	655
234	YasAsnArgMetTrpArgLysAsnAlaGlySerPheTyrAlaAsnAlaHisCys	255
657	ATCGGAAACAGACCTGATATGCAATTTGTCTCCAAACACATGCTGTGAGA	706
251	IleGlyTyrThrAspLeuAsnArgAsnPheAlaSerTyrHisTrpCysGlnG	267
707	AGATGATCATCAGTTCCTCATGCTCGGAAACATCAATGTCGACTTATATCTG	755
267	UGlyAlaIleSerSerSerCysSerGlnThrTyrCysGlyLeuTyrProG	284

```

757 ATCTGCAACCAAGGAGGAGGAGCGTGGCTAGTTCTTGAGAAAGAAATATC 806
|||||
284 IuSeGluProGluValIlySalValAlaSerPheLysAlaGluAsnIle 300
|||||
807 AACCAAGATTAAAGCATATCATCATGATTCATTCATCTCCAGCATATAGT 856
|||||
301 AaGlnIleLysAlaIlyIleSerMetHisSerTyrSerGlnHisIleVal 317
|||||
857 GTTTCATATTCCTCTATACACGAAGTAAGACAAAGACCATGAGAACTGT 906
|||||
317 IPherProTyrSerTyrThrArgSerLysSerLysAspHisGlnLueus 934
|||||
907 CHTAGTACCACGATGAAGCGTGGCTGTTCGACAAACAGTAAATAAT 956
|||||
334 eLeuValAlaSerGluAlaValAlaArgAlaIleGluLysThrSerLysAsn 350
|||||
957 ACCAGGTAAACACATGGCCATGGCGTCCGAACCTTATACCTAGTCTCTGG 1006
|||||
351 ThrArgTyrThrHisGlnHisGlySerGluThrLeuTyrIleuAlaProGlu 367
|||||
1007 AGGTGGGACGATTTGGATCTATGATTTGGGCATCAATATATTCGTT.... 1051H
|||||
367 yAllyGlyAspAspTyrPheTyrAspLeuGlyIleLysTyrSerPheThrI 384
|||||
1052 .....TAC 1055
|||||
384 IeGluLeuArgAspThrArgIlyThrIlyGlyPheLeuLeuProGluArgTyr 400
|||||
1055 ATCAAAACCCACCGCTGTAGAGAAAGCTTTGGCGGTCTCTTAATAATAGCTG 1104
|||||
401 IleLysProThrCysArgGluAlaPheAlaIleValSerLysIleAlaIrr 417
|||||
1105 GCATGTCATTAGGAATGTT 1123
|||||
417 pHisValIleArgAsnVal 423
|||||

seq_name: /cogn2_6/ptodate/1/1aa/5A_COMB.pep:US-08-430-787A-3

seq_documentation_block:
: Sequence 3, Application US/08430787A
: Patent No. 5593674
:
: GENERAL INFORMATION:
: APPLICANT: Drayna, Dennis T., Eaton, Dan L.
: TITLE OF INVENTION: No. 5593674e1 Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/430,787A
: FILING DATE: 27-APR-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/277,540
: FILING DATE: 19-JUL-1994
: APPLICATION NUMBER: 08/167727
: FILING DATE: 15-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/959944
: FILING DATE: 14-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/649591
: FILING DATE: 01-FEB-91
: ATTORNEY/AGENT INFORMATION:

```

NAME: Hasak, Janet E.
 REGISTRATION NUMBER: 28,616
 REFERENCE/DOCKET NUMBER: 689D1C1D1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-1896
 TELEFAX: 415/952-9881
 TELETYPE: 910/371-7168
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 423 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-430-787A-3

alignment_scores:

Quality: 1885.00 Length: 423
 Ratio: 5.122 Gaps: 2
 Percent Similarity: 86.998 Percent Identity: 86.525

alignment_block:

US-09-980-881-1 x US-08-430-787A-3

Align seg 1/1 to: US-08-430-787A-3 from: 1 to: 423

```

18 ATGAAGCTTTGCAAGCTTCGAGTCTGTACCCAAATGTTCTCTGTGTA 67
   |||||||
1  MetLysLeuYsSerLeuAlaValAlaValAlaProIlleValLeuPheCysGl 17
68 GCAGCATGTCTTGGCGTTCGAGAGTGGCCAAATTTCTAGCTGCTCTTCCTTA 117
   |||||||
17 uGlnHsValAlaPheAlaPheGlnSerGlnValAlaLeuAlaAlaLeuProA 34
118 GAACCTCTAGGCAAGTTCAGATTCTACAGAACTTACTACACATATAG 167
   |||||||
34 rghrSerArGlnValGlnValLeuGlnAsnLeuThrThrThrTyGln 50
168 ATGTGTTCTCGGCAAGCCGGAACAGCTGACCTTATTGTGAGAAAAACA 217
   |||||||
51 lIeValLeuTrpGlnProValAlaThrAlaAspLeuIlleValYsLysGln 67
218 AGTCAATTTTTTTGTAATGCATCTGATGTCGACAAATGTGAAGCCCAT 267
   |||||||
67 nValHsPhePheValAlaSnAlaSerAspValAlaAspValAlaGlnAsp 100
268 TAAATGTGACGGGAATTCATGCAGTGTCTTGGCGAGACGTGAAGAT 317
   |||||||
84 euAsnValSerGlnLeuProCysSerValLeuLeuAlaAspValGlnAsp 100
318 CTATATCAACAGACAGATTTCGAACGACAGTACGCCCGGAGCCCTCCG 367
   |||||||
101 LeuIlleGlnGlnIlleSerAsnAspThrValSerProAlaAlaSerAl 117
368 ATCGTACTATGAACAGTATCACTCACTAAATGAATCTATCTTGATAG 417
   |||||||
117 AserTyTyTyGlnGlnTyThrHsSerLeuAsnGlnIlleTySerTrpIlleG 134
418 AATTATTAACAGAGGAGCATCTGATATGCTTACAAAAATCCACATTTGA 467
   |||||||
134 lUrPheIlleThrGlnAlaGlnHsProAspMetLeuThrLysIlleHsIlleGln 150
468 TCCTCATTTTGAGAGAACCCACACTATGTTTAAAGTTTCTGAAAAAGA 517
   |||||||
151 SerSerPheGlnLysTyTrProLeuTyValAlaLeuLysValSerGlnLysGln 167
518 ACAAAACAGCCAAAATGCATATGATGATGATGATGATGATGATGATGATG 567
   |||||||
167 uGlnThrAlaLysAsnAlaIlleTrpLysPyrGlnLysAlaAlaArg 184
568 AATGATCTCTCTGCTTCTGCTGTGTGTTGATATAGGACAT..... 608
   |||||||
184 lUrTrpLysProAlaPheCysLeuTrpPheIlleGlnHsIlleThrGln 200

```

```

608 ..... 608
201 PheTyGlnIlleIlleGlnTyThrAsnLeuLeuArgLeuValAsp 217
608 ..... 608
217 eTyValMetProValAlaAsnValAlaSerGlnTyAspTySerTrpLysL 234
609 ..ATCGAATGTGGAGAAAGACCGTCTTCTGTGGACAAATCATATGC 656
   |||||||
234 yAsnAlaGMetTrpAlaLysAsnArgSerPheTyAlaAsnHsLysCys 250
657 ATCGAAGACAGCTGAATAGCACTTGTCTCCAAACAGCTGTGTGAGA 706
   |||||||
251 lIleGlnThrAspLeuAsnArgAsnPheAlaSerLysHsTrpCysGln 267
707 AGTGCATTCAGTTCTCATGCTCGGAACCTACTGTGACTTATTCCTG 756
   |||||||
267 uGlnAlaSerSerSerSerSerGlnTrpTyCysGlnLeuTyProG 284
757 AGTCAGAACAGAAAGTGAAGGAGTGGCTAGTTCTTGAGAGAAATATC 806
   |||||||
284 lUserGlnProGlnValAlaValAlaAlaSerPheLeuAlaArgAsnIlle 300
807 AACCAATTAAGCATACATCAGCATCATTCATCTCCAGCATATAGT 856
   |||||||
301 AsnGlnIlleLysAlaTyrlleSerMetHsSerTySerGlnHsIlleVa 317
857 GTTTCATATTTCTATACAGAAATTAACAAAGACATGAGAACTGT 906
   |||||||
317 lPheProTySerTyThrArgSerLysSerLysAspHsLysGlnGlnLeu 334
907 CTCTAGTAGCAGTGAAGAGTTCGTCTATGTGACAAATAGTAAATAT 956
   |||||||
334 eLeuValAlaSerGlnAlaValAlaArgAlaIlleGlnLysThrSerLysAsn 350
957 ACCAGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGTG 1006
   |||||||
351 ThrArgTyThrHsGlnHsGlnSerGlnThrLeuTyLeuAlaProGln 367
1007 AGGTGGGAGCATTTGATCTATGATTTGGCATCAATATTCCTT..... 1051
   |||||||
367 yGlnGlnAspAspTrpLysTrpLysPheGlnLysTySerPheThr 384
1052 .....TAC 1054
384 lLeuLeuAlaArgAspThrGlnTyThrTyGlnPheLeuLeuProGlnAlaArgTy 400
   |||
1055 ATCAAAACCACTGTAGAGAGCTTTGGCGTGTCTTAAATAGCTTG 1104
   |||||||
401 lIleLysProThrCysArgGlnAlaPheAlaValAlaValSerLysIlleAlaTr 417
1105 GCATGTCAATTAGGAATGTT 1123
   |||||||
417 PHisValIlleArgAsnVal 423

```

seq.name: /cgn2-6/plodata/1/1aa/5B-COMB.pep:us-08-869-057-2

seq documentation_block:

Sequence 2, Application US/08869057

Patent No. 5985562

GENERAL INFORMATION:

APPLICANT: Morser, Michael J

APPLICANT: Nagashima, Mariko

TITLE OF INVENTION: Method of Detecting Thrombotic Disease

TITLE OF INVENTION: Risk

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Berlex Biosciences Legal Department

STREET: 15049 San Pablo Avenue

CITY: Richmond

STATE: California

COUNTRY: USA


```

seq_name: /cgn2_6/plodata/1/1aa/5A.COMB.pep:us-07-649-591B-7
seq_documentation_block:
; Sequence 7, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
US-07-649-591B-7

alignment_scores:
Quality: 655.00 Length: 418
Ratio: 2.472 Gaps: 9
Percent Similarity: 63.397 Percent Identity: 35.885

alignment_block:
US-09-980-881-1 x US-07-649-591B-7 ..
Align seg 1/1 to: US-07-649-591B-7 from: 1 to: 417

33 CTTCAGATCCCTGTACCCATGTCTCTGTGAGCAGCATGTCTGCG 82
:::|||||::: ::::: |||
7 MetAlaValIleTyrThrIleuAlaIleAlaProValHis..... 20
83 GTTCCAGAGTGGCCAAAGTTCTAGTGGCTTCTTCCAGAACCTTAGGCAAG 132
|||||::: ::::: ::::: |||
21 PheAspArgGluValPheArgValIleuGlnAsnGluLysHisA 37
133 TTCCAGTCTTACAGATCTTACTACAACTATAGAGATTCTTCTGCGGAG 182
:::|||||:::|||||::: ::::: |||
37 IAsnValIleuLysAsnLeuThrGlnSerIleGluLeuAspPheTyr 53
183 CCGGTACAGAGTGCACCTTATTTGTGAAGAAACAAAGTCCATTTTTGT 232
||| ::::: |||::: ||| ||| |||
54 ProAspAlaIleHisAspIleAlaValAsnMetThrValAspPheArg 70
233 AAATGCATCTGATGTGCAACATGTGAAAGCCCATTTAAATGTGAGCGGAA 282
::: ::::: ::::: |||:::
70 IserGluLysGlnSerGlnThrIleGlnSerThrLeuGlnHisLysI 87

```

```

283 TTCCATGCAAGTCTCTGCGAGACGTGGAAGATCTTATTCACAGCAG 332
|| :::::||||: |||:::||||: |||:::||||:
87 LeHisTyrGluIleLeuIleHisAspLeuGlnGluIleGluLysGln 103
333 ATTTCC...AACGACAGATCAGCCCCGAGCTCCGACATGATATGCA 379
::: ::::: ||| ::::: |||
104 PheAspValLysAspGluIleAlaGlyArgHisSer.....TyrAl 117
380 ACAGTATCACTACTAATGAAATCTATCTTGATAGATTAATATACAG 429
:::|||||::: ::::: ||| ||| |||
117 AluSerTyrAsnAspThrAspLysIleValSerThrGluLysMetLeu 134
430 AGAGGATCTGATATGCTTACAAATAATCCATTTGATCTCATTTAG 479
||:|||||:::|||||:::||||: ||| ||| ||| |||
134 IuLysHisProGluMetValSerArgIleLysIleGlySerThrValGlu 150
480 AAGTACCATCTATATGTTTAAAGTTTCGGAAGAACAAACAGCCAA 529
||:|||||:::|||||:::||||: ::
151 AspAsnProLeuTyrValLeuLysIle...GlyLysLysAspGlyLys 166
530 AAATGCATATGATTCAGTGTGGAATCCATGCCAGATGGATGCTTTC 579
:::|||||:::|||||:::||||: ||| ||| ||| ||| |||
166 GlySalAlaIlePheMetAspCysGlyIleHisAlaArgGluTrpIleSer 183
580 CTGCTTTCTGCTTGTTGTTGATA.....GGCCAT 608
||||| ||| ||| ||| |||
183 roAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGluLys 199
609 AAT..... 611
200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeu 216
612 .....CGAATG 618
216 oValPheAsnValAspGlyTyrIleTrpSerThrPrlGlnAspArgMet 233
619 GGAAGAAAGAACGCTTTCTATGCAACATTCATTCAGGACAGAC 668
||||| ||| ||| ||| ||| ||| ||| ||| |||
233 rPArgLysAsnArgSerArgAsnGlnAsnSerThrCysIleGlyThrAsp 249
669 CTGGAATAGCAACTTGTCTCCAAACACGTGTGAGAGAGGTGATCCAG 718
||||| ||| ||| ||| |||
250 LeuAsnArgAsnPhe...AspValSerThrAspSerSerProAsnThr 265
719 TTCCATCATGCTCGAAGCTTACTGTGACTTATCTGAGTACGAGACAG 768
::: ||| ::::: ||| ||| ||| ||| ||| |||
265 nLysProCysLeuAsnValTyrArgGlyProAlaProGlnSerGluLys 282
769 AAGTGAAGCAGTGGCTAGTTCTTGAGAAGAAATATCAACAGATTAA 818
||:|||||:::|||||:::||||: ||| ||| ||| ||| |||
282 IuThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298
819 GCATACATCAGCATGATCATCTATCCAGCATATATGTTTCCATATTC 868
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 AlatYrIleThrPheHisSerTyrSerGlnMetLeuIleProTyrGln 315
869 CTATACACAGAGTAAAGCAAGACCATGAGAACTGTCTAGTAGGCA 918
||| ||| ::::: ||| ||| ||| ||| ||| ||| |||
315 YTrIleThrPheLysLeuProProAsnHisGlnAspLeuLysValAla 332
919 GTGAAGCAGTGTGCTATTTGACAAACTAGTAAATATCAACAGATTACA 968
||:|||||:::|||||:::||||: ||| ||| ||| ||| |||
332 rGluAlaIleThrAspAlaLeu...SerThrArgTyrGluThrArgTyrIle 347
969 CATGGCAATGGCTCAGAAACCTTATACACAGTCTCTGAGAGTGGGAGCA 1018
::: ||| ::::: ||| ||| ||| ||| ||| ||| ||| |||
348 TyrGlyProIleAlaSerThrIleTyrLysThrSerGlySerSerLeu 364
1019 TTGATCTATGATTTGGCATCAATATTCGTTTAC..... 1054
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 pTrpValTyrAspLeuGlyIleLysHisIleThrPheAlaPheGluLeuAla 381

```

```

1055 .....ATCAACCCACC 1066
381 splysglylysSerglyPheleuleuProgluSerArglyleLysProthr 397
1067 TGTAGAACCTTTTGGCGTGTCTTAATACCTTGGCATGTCATTAG 1116
398 CyalysleuThrMetLeuSerValLysPheleleLyslyleleuLys 414
1117 GAAT 1120
1118 :|||
414 SASN 415

seq_name: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:us-08-277-540-7

seq_documentation_block:
; Sequence 7, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patln (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277, 540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-277-540-7

alignment_scores:
Quality: 655.00 Length: 418
Ratio: 2.472 Gaps: 9
Percent Similarity: 63.397 Percent Identity: 35.885

alignment_block:
US-09-980-881-1 x US-08-277-540-7 ..
Align seg 1/1 to: US-08-277-540-7 from: 1 to: 417
33 CTTCAGATCCTGTACCATGTGTCTCTCTGTAGAGCAGCATGCTTCCG 82

```

```

7 MetAlaValIleYrThrThrLeuAlaIleAlaProValHis..... 20
8 GTTCAGAGATGGCCAGCTTCTAGCTGCTCTCTAGAACCTTAGGCAAG 132
21 PheAspArgIleValValPheArgValLysLeuGlnAsnGluLysHisA 37
133 TTCAAGTTCACAGAAATCTTACTACACATATGAGATTGTTCTGCGAG 182
183 CCGGTACAGCTGACCTTATTTGTGAAGAAAAACAAAGCTCATTTTTGT 232
54 ProAspAlaIleHisAspIleAlaValAsnMetThrValAspPheArgVa 70
233 AATGATCATGTATGTCGACAAATGTGAAGCCCATTTTAATGAGCGGGA 282
70 LserGluLysGluSerGlnThrIleGlnSerThrLeuGlnGlnHisLysI 87
283 TTCCATGCACTGTCTGTGCGACAGCTGGAAGATCTTATTCACAGCAG 332
87 LHisTyrGluIleLeuIleHisAspLeuGlnGlnGluIleGluLysGln 103
333 ATTGCC...AACGACACAGTCAGCCCCGAGCTCCGATGCTACTATGA 379
104 PheAspValLysAspGluIleAlaGlyArgHisSer.....TyrAl 117
380 ACAGATGACCTGACCTAAATGAATCTATTTGATAGATTTTAACTG 429
117 AluTyrAsnAspIlePheValLysIleValSerThrGluLysMetLeuG 134
430 AGAGCATCTGATATGCTTACCAAAATCCATGATGATCTCATTTGAG 479
134 LuLysHisProGluMetValSerArgIleLysIleGlySerThrValGlu 150
480 AAGTACCCACTGTATGTTTAAAGTTCTGGAAGAACAAACAGCCAA 529
151 AspAsnProLeuTyrValLeuLysIle...GlyLysAspGlyGluArg 166
530 AAATGCCATGATGATGACCTGGAATCCATGCCAGAAATGATCTGTC 579
166 GlyAlaIlePheMetAspCysGlyIleHisAlaArgGluThrIleSerP 183
580 CTGCTTCTGCTGTGTTGATCA.....GGCCAT 608
183 roAlaPheCysGlnThrPheValTyrGlnAlaThrLysSerTyrCylLys 199
609 AAT..... 611
200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuPr 216
612 .....CGAAGT 618
216 oValPheAsnValAspGlyTyrIleThrSerThrThrGlnAspArgMetT 233
619 GGAGAAAGAACCTTCTTCTATGCGAACAATCATGATGATGCGAAGCAG 668
233 TrpArgLysAsnArgSerArgAsnGlnAsnSerThrCysIleGlyThrAsp 249
669 CTGAATACCACTTTGTTCCAAACACTGCTGTAGAGAGGTGCATCCAG 718
250 LeuAsnArgAsnPhe...AspValSerThrPaspSerSerProAsnThrAs 265
719 TTCCATGCTGCGAAGCTTACTGTCGACTTATGCTTATGTCAGAACAGC 768
265 nLysProCysLeuAsnValTyrArgGlyProAlaProIleSerGluLysG 282
769 AAGTGAAGCAGATGCTAGTTCTTGAGAAAGAAATATCAACAGATTAA 818
282 LuThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298
819 GCATACATCAGCATGATTCATACCTCCAGCATATAGTGTTCATATTC 868

```

```

299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrG1 315
869 CTAATACAGCAAGTAAAGCAAGACCATGAGAACTGTCTTAGTACCA 918
      |||||      :|||:|||||      |||||
315 yTyrThrPheLeuProGlnHisGlnAspLeuLeuValAlaIaa 332
      |||||      :|||:|||||      |||||
919 GTCAACGACGTGCTGTATGACAAAAGTGTAAATACCAGGTATACA 968
      |||||      :|||:|||||      |||||
332 rGleAlaThrAspAlaLeu...SerThrArgTyrGlnThrArgTyrIle 347
969 CATGCCATGCGTCAGAAACCTTATCTGACCTGAGAGGGAGCA 1018
      :|||:|||||:|||||:|||||      :|||:|||||:
348 TyrcylProIleAlaSerThrIleTyrLysThrSerGlySerSerLeuAs 364
1019 TTGGATCTATGATTTGGCATCAATATTCGTTAC..... 1054
      |||||:|||||:|||||:|||||:|||||
364 prpValTyrAspLeuGlyIleLysHisThrPheAlaPheGlnLeuArg 381
1055 .....ATCAACCCACC 1066
381 spLysGlyLysSerGlyPheLeuLeuProGlnSerArgIleLysProThr 397
1067 TGTAGAGAACTTTGCGCGTCTCTTAAATAAGCTTGGCATGTCAATAG 1116
      |||||:|||||:|||||:|||||      |||||
398 CysLysGlnThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLy 414
1117 GAAT 1120
      :|||
414 sAsn 415

seq_name: /cgn2_6/plodata/1/1aa/5A-COMB.pep:US-08-430-787A-7

seq_documentation_block:
  Sequence 7, Application US/08430787A
  Patent No. 5593674
  GENERAL INFORMATION:
    APPLICANT: Drayna, Dennis T., Eaton, Dan L.
    TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
    NUMBER OF SEQUENCES: 8
    CORRESPONDENCE ADDRESS:
      ADDRESS: Genentech, Inc.
      STREET: 460 Point San Bruno Blvd
      CITY: South San Francisco
      STATE: California
      COUNTRY: USA
      ZIP: 94080
  COMPUTER READABLE FORM:
    MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patin (Genentech)
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/430,787A
    FILING DATE: 27-APR-1995
    CLASSIFICATION: 514
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/277,540
    FILING DATE: 19-JUL-1994
    APPLICATION NUMBER: 08/167727
    FILING DATE: 15-DEC-1993
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 07/959944
      FILING DATE: 14-OCT-1992
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 07/649591
      FILING DATE: 01-FEB-91
  ATTORNEY/AGENT INFORMATION:
    NAME: Haack, Janet E. 28,616
    REGISTRATION NUMBER:
    REFERENCE/DOCKET NUMBER: 689D1C1D1
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 415/225-1896
    TELEFAX: 415/952-9881

```

```

;
; TELEX: 910/371-7168
;
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 417 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;
US-08-430-787A-7

alignment_scores:
  Quality: 655.00      Length: 418
  Ratio: 2.472          Gaps: 9
  Percent Similarity: 63.397      Percent Identity: 35.885

alignment_block:
US-09-980-881-1 x US-08-430-787A-7 ..

Align seq 1/1 to: US-08-430-787A-7 from: 1 to: 417

33 CTTCAGTCCTGTGACCCATGTTCTCTCTGTGAGCAGCATGCTTGGC 82
   :|||:|||||:      :|||:|||||:
7 MetAlaValIleTyrThrThrIleAlaIleAlaProValHis..... 20
83 GTTCCAGAGTGGCCAAAGTTTACAGTGTCTTCTTCTTACAGCCTAGGCAAG 132
   |||||:|||||:|||||:|||||:|||||
21 .PheAspArgGlnLysValAlaPheArgValLysLeuGlnAsnGlnLysHisA 37
133 TTCAAGTTTACAGAAATCTTACTATACAAACATATGAGATGTTCTCTGGCAG 182
   :|||:|||||:|||||:|||||:|||||
37 IAservAlleuLysAsnLeuThrGlnSerIleGlnLeuAspPheThrTyr 53
183 CCGGTAAACAGCTGACCTTATGTGAAGAAAACAAAGTCCATTTTGT 232
   |||:|||||:|||||:|||||:|||||
54 ProAspAlaIleHisAspIleAlaValAlaMetThrValAspPheArg 70
233 AAATGCACTGATGTGCAGCAATGTGAAGCCCATTTAAATGTAGCGGAA 282
   |||||:|||||:|||||:|||||:|||||
70 IserGlnLysGlnSerGlnThrIleGlnSerThrLeuGlnLysI 87
283 TTCCATGACGTGCTTGTGTCGACAGTGGAGATGTTATTAACAGCAG 332
   ||:|||||:|||||:|||||:|||||
87 LeHisTyrGlnIleLeuIleHisAspLeuGlnGlnIleGlnLysGln 103
333 ATTTC...AACGACAGCTGACCCCGAGCTCCGACATGCTATGCA 379
   :|||:|||||:|||||:|||||
104 PheAspValLysAspGlnIleAlaGlyArgHisSer.....TyrAl 117
380 ACAATGACATGCTAATGAAATGATGTTCTGTGATAGAAATTTATACAG 429
   :|||:|||||:|||||:|||||
117 AluSTyrAsnAspTrpAspLysIleValSerTrpGlnLysMetLeuG 134
430 AGAGCATCTGATATGCTTACAAAATCCAAATGATGATGCTATTTGAG 479
   ||:|||||:|||||:|||||:|||||
134 IuLysHisProGlnMetValSerArgIleLysIleLysThrValGln 150
480 AAGTACCCACTATGTTTAAAGTTTGTGAAAAGAACAAACAGCCAA 529
   |||||:|||||:|||||:|||||
151 AspAsnProLeuTyrValLeuLysIle...GlyLysLysAspArgLysL 166
530 AATGGCATATGATGATGATGATGATGATGATGATGATGATGATGATG 579
   :|||:|||||:|||||:|||||:|||||
166 GlyLysAlaIlePheMetAspCysGlyIleHisAlaArgGlnTrpL 183
580 CTGCTTCTGCTGTGGTTCATA.....GGCAT 608
   |||||:|||||:|||||:|||||
183 roAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199
609 AAT..... 611
   |||
200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuPr 216
612 .....CGAATGT 618
   |||||

```



```

216 ovalpheasnvalaspelytyrlletpsertrpThglnaspargmect 233
619 GGAGAAAGAACGGTCTTCTTATGACGAACATTCATTCGATCGAAGAC 668
233 rpatgylasnarargserArgasnlnasnsertlrcysllleclYthasp 249
669 CTGAATGCAACTTTCCTCCAAACACTGCTGAGAGAGGATGCATCCAG 718
250 leuasnArgasnph...AspvalsetrtpaspserserProasnthas 265
719 TTCCTCATGCGGGAACCTACTGACCTTATTCCTGAGCAGAACACG 768
265 nlvsprocysleuasnvaltyrarglyproalaproglubserglulysg 282
769 AAGTGAAGCAGTGGCTAGTATTTCTTGAGAGAAATATCAACAGATTAA 818
282 luthrlysalavalThrasnphelleargserHlsleasnsertlleys 298
819 GCATACATCAGCATGATCTACTCCAGCATATGAGTGTTCATATTC 868
299 Alatyrlletlrphelnsertyserserlmetleuileuileprotygl 315
869 CTATACAGAGTAAAGCAAGACATGAGGACGTCTCTAGTACGCA 918
315 YTYTThrphelysleupropProasnlhslaspleuulysvalalaa 332
919 GTGAGCAGTCTGCTATTTGACAAACTAGTAAATATCCAGATATACA 968
332 rgllealathraspaleu...SerthrarglyrgluthrArgtyrlle 347
969 CATGGCCATGGCTGCAAAACCTTATACCTAGCTCTGAGGTGGGAGCA 1018
348 TyrclyprolealaserThriletyrlysergllyserSerleuas 364
1019 TTGATCATGATTTGGGATCAATATATGCTTAC..... 1054
364 prtrpvaltyraspleuuglylelyshlsthphelalphecluleuarg 381
1055 .....ATCAAAACCCACC 1066
381 splysglylyserglypheluleuuprogIuserArglllelyserprothr 397
1067 TGTAGACAGCTTTTGGCGCTGTCTCTTAAATAGCTTGCAATGCATTA 1116
398 CyslysglutlrmelTleuserVallyspheillealalystrylleu 414
1117 GAAT 1120
414 sasn 415
seq_name: /cgn2_6/prodata/1/laa/5B_COMB.pep:US-08-860-882A-57
seq_documentation block:
; Sequence 57, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGEITE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUM
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.

```

```

; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-860-882A-57

alignment_scores:
Quality: 651.50 Length: 419
Ratio: 2.404 Gaps: 8
Percent Similarity: 64.678 Percent Identity: 33.890

alignment block:
US-09-980-881-1 x US-08-860-882A-57 ..
Align seg 1/1 to: US-08-860-882A-57 from: 1 to: 415

33 CTTCGAGTCTGTGACCCATTTCTCTCTGTGAGCAGATGCTTGGC 82
2 leuileuValleuValThrValalaleuAlaserlAlhshlslygl 18
83 G...TTCCAGAGTGGCCAAAGTTTCTAGCTCTCTTGAACCTCTAGCC 129
18 whlspheluglyglulysValrpheargValasnlalulaspgrluasn 35
130 AAGTCAAGTCTACAGATCTTACTACAGATATGAGATGTCTGCG 179
35 lsleasnllleleargluleuAlaserThrThglnlleasprhetrp 51
180 CAGCCGTAACAGCTGACCTTATTTGTGAAGAAAACAGTCCATTTT 229
52 lypProaspservallThrglnllelysprohlssetrlhValasprheAr 68
230 TGTAAATGATCTGATGTGACATATGTGAAGCCCATTTAAATGTGAGCG 279
68 yAllysalagluAsprThrValThrValglulasnValleuylsglnaang 85
280 GAATTCATGACAGTGTCTGTGCGACAGCGTGAAGATCTTATTAACG 329
85 lulenglnltyrlysalvalleuileserasnleuargasnvalalglula 101
330 CAGATTCCACAGCAGATCAGCCCGAGCCCTCGCATCTACTATGA 379
102 glnphleaspservArgVal.....Argalathrglyhlssetlygl 115
380 ACAGTATCACTCAGTAATGAATATATTTCTGATAGATTTAATACTG 429
115 ulystyrasnlystrpgluthrlleaglualatrpHrglnglnvalalAt 132
430 AGAGCATCCGATATGCTTACAAAATCCAAATCCAGATCCCATTTGG 479
132 hrgluasnProalaleuileserArgservallleclYthtrphelgu 148

```



```

306 GACGTGGAAGATCTTATTCACAGCATTTCC...AAGCAGACAGTCAG 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 Aspleuengingluilegyluslphesapvalylsglualsplel 111
353 CCCCCGAGCCCGATGCTACTATGACATACACTGACATAAAGAA 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 oglArlghliser.....Tyralalyslyrasmansntrpglulysl 125
403 TCTATTTCTGGATAGATTATTAAGTGAAGCATCTGTATATGCTTCA 452
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 levalAlatrptlhrGluysmethetalsplyslyrproglmetlvalser 141
453 AAAATCCATGGATGCTCATTTGAGAGATGACCACTATATGTTTAA 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 ArgileylsilegylserThrValglusapsnproleuTyralleuyl 158
503 GGTTCGTGGAAGAACAAACAGCCAAAATGSCATATGATGATGCTGTG 552
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 sile...glYglulysasnglualrgralysalallephemetalspcysg 174
553 GAATGCATGCCAGAGAAATGATGATCTCTGCTTCTGCTGTGTTCA 602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 LytlelnlsAlatrglultrValserproAlarhescysgltrpHeval 190
602 ..... 602
191 TyrglnAlatrhlYsThrTyrglyArgasnlyslmetThrlyleu 207
603 .....GCCCAT. 608
207 uasprargmetasnphetyrilleuProvalphesalnValaspllyrYrI 224
609 .....AATCGAATGTGAGAGAAACCGTTCTTCTAT 641
224 letprsertrptlhrlyasnargmettrparglyasnargserlyasn 240
642 GCGAACAATCTTGCATCGGAGACAGACCTGAATAGCAATTTGCTCCA 691
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 GlnasnserlyscysilegylThrAspleuasnArgasnph...Asnal 256
692 ACACGTGATGAGGAGGTGCATGCTCATGCTGCGAAGAACTCT 741
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 aserTrpanserlIerproasnThrAsnAspTocysAlaspsnlyrTA 273
742 GTGACATTTATCTGAGTCAGACACAGAGTAAGAGCGAGTGTAGTTTC 791
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
273 rglYserAlarproglusserglulysglulhrlysalalValThrasnph 289
792 TTGAGAAATATATCAACAGATTAAGCATATACATGACATTCATTA 841
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 IleargserlnsleuasnglulileysalalYrtleThrlnhisserly 306
842 CTCACGATATAGTGTTCATATTCATATACAGAAAGTAAAGCAAG 891
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 rSerGlmetleuLeupheproTyrglyrlyThrSerlyleuProProA 323
892 ACCATGAGGAATCTCTAGTACAGAGTGAAGAGTCTGCTATTTGAC 941
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 snhlsgluAspleualalysValalalysilegylThrAspvalleu... 338
942 AAAACTGTATAAATACAGGATATACAGATGCGCTGAGAAACCTT 991
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 SerThrarglyrGluThrArglyrIleTyrglyProillegluserThrI 355
992 ATACCTAGCTCTGAGGTGGAGAGATGATGATATGATTTGGGACATCA 1041
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
355 etYrProilleserIyserSerleuAspTrpalatylAspleuyllel 372
1042 AATATTCGTTTAC..... 1054
372 yshlstrlnrphelAlarhegluldeuArgasplysglylsrpheglYrheleu 388

```

```

1055 .....ATCAAAACCACTGTGTAGAGAACTTTTGGCGCTGT 1089
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 LeuprogluserArgilleysprohrtrcysargglulhrmetleuAlava 405
1090 CTCCTAAATAGCTTGGCATGCTCATAGGAAT 1120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 llysrphelAlalalyslyrIleleuylshls 415
seq_name: /cgn2_6/prodata/1/lae/5A_COWB pep:US-08-277-540-6
seq_documentation block:
? Sequence 6, Application US/08277540
? Patent No. 5474901
? GENERAL INFORMATION:
? APPLICANT: Dreyna, Dennis T., Eaton, Dan L.
? TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genentech, Inc.
? STREET: 460 Point San Bruno Blvd
? CITY: South San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94080
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: patin (Genentech)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/277,540
? FILING DATE: 19-JUL-1994
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/167727
? FILING DATE: 15-DEC-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 07/959944
? FILING DATE: 14-OCT-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 07/649591
? FILING DATE: 01-FEB-91
? ATTORNEY/AGENT INFORMATION:
? NAME: Hasak, Janet E.
? REGISTRATION NUMBER: 28,616
? REFERENCE/DOCKET NUMBER: 689D1C1D1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415/225-1896
? TELEFAX: 415/952-9881
? TELEX: 910/371-7168
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 417 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? US-08-277-540-6
alignment_scores:
Quality: 647.50 Length: 427
Ratio: 2.398 Gaps: 10
Percent Similarity: 63.232 Percent Identity: 33.489
alignment_block:
US-09-980-881-1 x US-08-277-540-6 ..
Align seg 1/1 to: US-08-277-540-6 from: 1 to: 417
6 AATGCTGTGGAGATGAGCTTTGACGCTTGAGACCTTGATACCAATGT 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 LeuprovalglyleuileAlatrThrleuAlalle...Alaproval.. 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 TCTCTCTGTGAGCAGCATGCTTCGCGTTCAGAGTGGCAAGTTCTAG 105

```

```

20 ..... ||||| :|||:
106 CTGCTCTCTAGAACCTTAGGCAAGTTCAAGTTCTACAGATCTTACT 155
    ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
28 rGValLysPrGlnAspGlnLysGlnAlaAspIleIleLysAspLeuAla 44
    ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
156 ACAACATATAGATGTTCTCTGCGACCCGGTAAACAGCTGACCTTATGT 205
    ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
45 LysThrSngLLeuAspPheTrpTyLProGlyAlaThrHisValAl 61
206 GAGGAAAAACAAGTCATTTTGTAAATGCATCTGATGTGCACATG 255
    :|||: ||||| ||||| :|||: ||||| :|||: ||||| :|||:
61 AlaAspMetLeuValAspPheArgValSerGlnLysGlnSerGlnAla 78
256 TGAAGCCCATTTAAATGTGAGCGGAATTCACATGCTGCTTCTGCA 305
    :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
78 LeGlnSerAlaLeuAspGlnAsnLysMetHisTyLrGlnIleLeuIleHis 94
306 GAGCTGAGAGATCTTATTCACACAGCATTTCC...AACGACACAGTCAG 352
    ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
95 AspLeuGlnLuglnIleGlnLysGlnPheAspValLysGlnAspIlePr 111
353 CCCCCGAGCTCGCATCTACTATGAACAGTACCTCACTCAATAATGAA 402
    ||||| ||||| ||||| :|||: ||||| :|||: ||||| :|||:
111 OGlyAlrGHisSer.....TyLAlaLysTyLAsnAsnTrpGlnLysI 125
403 TCTATTCTTGAGATATTTATACAGAGCGCATCTGATATGCTTACA 452
    ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
125 LeValAlaTrpThrGlnLysMetLeuAspLysTyLrProGlnMetValSer 141
453 AAATGCACATGATCTTATTCAGAGATACCCACTCATGTGTTTAA 502
    ||||| ||||| ||||| :|||: ||||| :|||: ||||| :|||:
142 ArgIleLysIleGlySerThrValGlnAspAsnProLeuTyLValLeuLy 158
503 GGTTCGAAAAAGAACAGCAAGCAAAATGCCATATGATGTTGACGTG 552
    ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
158 sIle...GlyGlnLysAsnGlnLysAlaIlePheMetLeuAspGly 174
553 GAATTCATGCCAGAGAAATGATCTCTCTCTCTCTCTCTCTCTCTCT 602
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
174 LylleHisAlaArgLysIleTrpValSerProAlaPheCysGlnTrpPheVal 190
602 .....
191 TyrGlnAlaThrLysThrTyLrGlyArgAsnLysIleMetThrLysLeu 207
603 .....GGCCAT. 608
207 uAspArgMetAsnPheTyLrIleLeuProValAlPheAsnValAspGlyTyL 224
609 .....AATCGAATGTGAGAAAGAACCGTCTCTTCTAT 641
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
224 LeTrpSerTrpThrLysAsnArgMetTrpArgLysAsnArgSerLysAsn 240
642 GCGAACAATATTCATTCGGAACAGACTGAATAGCAACTTGTCTGCCA 691
    ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
241 GlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnHe...AsnAl 256
692 ACACTGGTGTGAGGAAGGTGATCGATCTCATGCTCGAAGAACCTACT 741
    ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
256 AserTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyL 743
742 GTGGACTTTATCTGAGTCAAGAACAGAACTGAAGCAGTGGCTAGTTTC 791
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 rGlyLysSerAlaProGlnSerGlnLysGlnThrLysAlaValAlrAsnPhe 289
792 TTGAGAGAAATATCAACAGATTAAGCATACATCAGCATGATTCATA 841
    :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
290 IleArgSerHisLeuAsnGlnLysValTyLrIleThrPheHisSerTy 306
842 CTCACGATATATGTTGTTCCATATTCCTATACAGCAAGTAAAGCAAG 891
    ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:

```

```

306 rSerGlnMetLeuLeuPheProTyLrGlyTyLrHisSerLysLeuProPro 323
892 ACATAGAGAGACTGCTCTACTAGCACAGTGAAGCAGTTCCGCTATTGAC 941
    :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
323 snHisGlnAspLeuAlaLysValAlaLysIleGlyThrAspValLeu... 338
942 AAAACTAGTAAATAATACAGGTATACATAGCCATGCTCGAAGACCTT 991
    :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
339 SerThrArgTyLrGlnThrArgTyLrIleTyLrGlyProIleGlnSerThrI 355
992 ATACCTAGCTCTCTGAGTGGGAGCAGTGTGATATGATTTGGGCATCA 1041
    :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
355 eTyLProIleSerGlySerLeuAspTrpAlaTyLrAspLeuGlyIle 372
1042 AATATGCTTTAC..... 1054
372 yshIstThrPheAlaPheGlnLeuArgAspLysGlyLysPheGlyPheLeu 388
1055 .....ATCAAAACCCACCTGTAGAGACCTTTGGCCGCTG 1089
389 LeuProGlnSerArgIleLysProThrCysArgGlnThrMetLeuAlaVal 405
1090 CTCATAAATAGCTTGGCATGTGATTAGAAT 1120
    ||||| ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
405 LysPheIleAlaLysTyLrIleLeuLysHis 415
seq_name: /cgn2_6/ptodata/1/laa/5A_COMB.pep:us-08-430-787A-6
seq documentation block:
; Sequence 6, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patln (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids

```

TYPE: amino acid
TOPOLOGY: linear
US-08-430-787A-6

alignment_scores:
Quality: 647.50 Length: 427
Ratio: 2.398 Gaps: 10
Percent Similarity: 63.232 Percent Identity: 33.489

alignment_block:

US-09-980-881-1 x US-08-430-787A-6 ..

Align seq 1/1 to: US-08-430-787A-6 from: 1 to: 417

```

6  ATTCTGTGGATGAAGCTTTCAGCCCTTCAGTCCCTTACCCATTGT 55
   :: ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
5  LeuProValGlyLeuIleAlaThrThrLeuAlaIle..AlaProVal.. 19
   ||| ||||| ::||| ||| ::||| ::||| ::||| ::||| ::|||
56 TCCTCTCTGACGACGATGCTTCGCGTTCAGAGTGGCCAGTCTTAG 105
   ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
20  ....ArgPheAspArgGluValPheA 28
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
106 CTGCTCTTCAGACCTTCAGGCAAGTTCAGTTCACAGATCTTACT 155
   :: ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
28  rGValLysProGlnAspGluLysGlnAlaAspIleIleLysAspLeuAla 44
   ||| ||||| ::||| ||| ::||| ::||| ::||| ::||| ::|||
156 ACAATATGATGATTTGCTCTGACGCGGCTAACAGCTGACCTTAATGT 205
   ||| ||||| ::||| ||| ::||| ::||| ::||| ::||| ::|||
45  LysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisValAl 61
   ::||| ::||| ||| ||||| ::||| ::||| ::||| ::|||
206 GAAGAAAAACAGTCATTTTGTAAATGATCTGATGTCGACAATG 255
   ::||| ::||| ||| ||||| ::||| ::||| ::||| ::|||
61  aAlaAsnMetValAspPheArgValSerGluLysSerGlnAlaI 78
   ::||| ::||| ||| ||||| ::||| ::||| ::||| ::|||
256 TGAAGCCCATTAATGTGAGCGGCAATCCATGACAGTCTTGCTGCA 305
   ::||| ::||| ||| ||||| ::||| ::||| ::||| ::|||
78  LeuGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94
   ::||| ::||| ||| ||||| ::||| ::||| ::||| ::|||
306 GACGTGGAAGATCTTATTCACAGCAGATTTCC...AAGCAGACAGTCAG 352
   ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
95  AspLeuGlnGluIleGluLysGlnPheAspValLysGluAspIlePr 111
   ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
353 CCCCCGAGCTCCGCAATCGTATGAAACATGATCACTCACTAAATGAA 402
   ||||| ::||| ||| ||||| ::||| ::||| ::||| ::|||
111  oGlyArgHisser.....TyrAlaLysTyrAsnAsnTrpGluLysI 125
   ||||| ::||| ||| ||||| ::||| ::||| ::||| ::|||
403 TCCTATCTGATGATTTATTAAGTGAAGGCACTCCGATATGCTTACA 452
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
125  LeValAlaTrpThrGluLysMetMetAspLysTyrProGluMetValSer 141
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
453 AAATTCACATTTGATTCCTCATTTGAGAGTAAACCACTCTATGTTTAA 502
   ::||| ||||| ::||| ::||| ::||| ::||| ::||| ::|||
142  ArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuL 158
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
503 GGTTCCTGGAAGAAGACAAACAGCCAAAATGATATGATGATGACTGTG 552
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
158  sile...GlyGluLysAsnGluArgGlyAlaIlePheMetAspCysG 174
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
553 GAATTCATGCCAGAGAAATGATCTCTGCTTCTGCTGCTGCTGCTGCTA 602
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
174  LylIleHisAlaArgLutTrpValSerProAlaPheCysGlnTrpPheVal 190
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
602  .... 602
191  TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeu 207
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
603  ....GGCCAT. 608
207  uAspArgMetAsnPheTyrIleLeuProValPheAsnValAspGlyTyrI 224
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
609  ....AATGAAATGGAGAAAGAACCGTCTTTCAT 641
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||

```

```

224  LeTPSerTrpThrLysAsnArgMetTrpArgLysAsnArgSerLysAsn 240
642  GCGAACATCATTCATCGACGAAACAGACCTGATAGCACTTGTCTCCA 691
   ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
241  GlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPro...AsnAl 256
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
692  ACACGTGTGAGGAGAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACT 741
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
256  AserTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrA 273
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
742  GTGACCTTATCTGACGACGACGACGACGACGACGACGACGACGACGAC 791
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
273  rGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPro 289
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
792  TTGAGAGAGAAATATCACACGATTAAGCATATACATGATGATGATGAT 841
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
290  IleArgSerHisLeuAsnGluIleLysValLysIleThrPheHisSerT 306
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
842  CTCACGACATATAGTGTTCATTCCTATACACGAGTAAAGCAAAAG 891
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
306  rSerGlnMetLeuLeuPheProTyrGlyTyrThrSerLysLeuProProA 323
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
892  ACCATGAGAGACTGCTCTAGTACGACGACGACGACGACGACGACGAC 941
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
333  snHisGluAspLeuAlaLysValAlaLysIleGlyThrAspValLeu... 338
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
942  AAACTAGTAAATAATACCGAGTATACACATGCGCATGCGTCAAAACCTT 991
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
339  SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThrI 355
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
992  ATACCTAGCTCTCGGAGGTGGGACGATGATGATGATGATGATGATGAT 1041
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
355  eTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIle 372
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
1042  AATATGCTTAC..... 1054
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
372  yshisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
1055  ....ATCAACCCACCTGTAGAGAACCTTTTGCCGCTGT 1089
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
389  LeuProGluSerArgGlyLeuProThrCysArgGluThrMetLeuAlaVa 405
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
1090  CTCATAAATAGCTTGGCATGTCATTAGAAT 1120
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
405  LysPheIleAlaLysTyrIleLeuLysHis 415
   ||||| ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||

```

seq_name: /cgn2_6/protdata/1/laa/5a_COMB.pep:US-08-696-139-2

seq_documentation_block:

Sequence 2, Application US/08696139

Patent No. 5672496

GENERAL INFORMATION:

APPLICANT: Fayerman, Jeffrey T.

APPLICANT: Greenen, David P.

APPLICANT: Hersberger, Charles L.

APPLICANT: Larson, Jeffrey L.

APPLICANT: Steiner, Jane L.

APPLICANT: Zhang, Haichao

TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Eli Lilly and Company

STREET: Lilly Corporate Center

CITY: Indianapolis

STATE: Indiana

COUNTRY: United States of America

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

PCT-US96-00995-6

```
alignment_scores:
    Quality: 607.50      Length: 305
    Ratio: 3.148         Gaps: 6
Percent Similarity: 63.279   Percent Identity: 41.967
```

alignment_block: TS-09-980-881-1 x PCT-TS96-00995-6

Align seg 1/1 to: PCT-US96-00995-6 from: 1 to: 307

[illegible]

```

999  GTCCTGGAGCGTGGGAGACATGTGATCATGATGTTGGGCATCAAAATATC 1048
248  AATAAGTGTGYSERASPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 264
1049  GTTATCATC..... 1057
      |||||
264  rherhrrhpegluleauragspthrglypnebhelypbeleuleup 281
1058  .....AACCACACCTGTACAGAAAGCTTTTGCCGCTGTCTTAA 1096
      :: |||||:::|||||::: |||||:::
281  luserglmlleargslntfrchysgslnglulrmetleualavalalyslyr 297
1097  ATAGCTTGGCATGTC 1111
      ||||| ::|||
298  lleaalaasntyrval 302

```

This Page Blank (uspto)

117 aserYtYrYtGlnGlnTyrHisSerLeuasnGlnUleYrSerTrpIleG 134

418 AATTATTAACGTGAGAGCGCATCCGTGATGCTTACAAAAATCCACATTGGA 467

134 TrrpelleIerGlnUaGlnHisProAsnMetLeuThrYsIleHisIleGly 150

468 TCCCATTTGAGAAGTACCCACATCTCTGTTTAAAGCTTCGGAAATAA 517

151 SerSerPheUuYrYsYrProleuYrValIleuYsValSerGlnYsGcl 167

518 ACAAAACAGCCAAAATGCCATATGATGATTACTGTGGATCCATGCCAGAG 567

167 uGlnTrpAlaYsAsnAlaIleerTrpIleAspYsGclIleHisAlaAsG 184

568 AATGAACTCTCCGCTTCTGCTGTGGTGCATACCCCAT..... 608

184 TrrTrpIleSerProAlaPheCysUleTrpPheIleYsIleHisIleThrGln 200

608 608

201 PheYrGlyIleIleGlyGlnYrThrAsnLeuLeuArgLeuValAspH 217

608 608

217 eYrYrValMetProValValAsnValAspGlyYrAspYrSerTrpYsL 234

609 . AATGCAATGTGGAGAAAGAACCCGTTCTTTTATGCGCAACATCATTCG 656

234 YAsnArGmetTrpArgYsAsnAlaGserPheYrAlaAsnHisIleYs 250

657 ATGGGAACAGACSTGAATAGCAATTGTGCTGCCAAACACATGCTGGAGA 706

251 TleGlyThrAspLeuAsnArGAsnArPheAlaSerYsHisIleTrpCysGln 267

707 AGGTGATCCAGTTCCTCATGCTCGGAAACCTACAGTGGACTTTCCTG 756

267 uGlnYrAlaSerSerSerSerCysSerGlnThrYrYsGlyUleYrProG 284

757 AGTCAGAACCCAGACTGAAGGACAGTGGCTAGTTCTTGAGAAATAATTC 806

284 InsertUrProGlnUalYsAlaValAlaSerPheLeuArgYrAsnIle 300

807 AACCAATTAAGCACTACATCAAGATGCATTCATCTCCAGCAATATGT 856

301 AsnGlnIleuYrAlaYrIleSerMetHisSerYrSerGlnHisIleVa 317

857 GTTTCATATTCCTATACAGCAAGTAAAGCAAAAGACATGAGCAACTG 906

317 TrpProYrYrSerYrThrArgSerYsSerYsArpHisGlnUleuS 334

907 CTTCTAGTACCCAGTGAAGAGCTTCCTGCTATTTGACAAACATGAAAT 956

334 eTrpUuValAlaSerGlnUalValAlaGlnAlaIleGlnUyAsH SerYsAsn 350

957 ACCAGGTATATACATAGGCGCATGGCTGCAGAAACCTATACSTAGTCCG 1006

351 TrpArgYrThrHisGlnHisIleUerGclUthrIleuYrTrleuAlaProGcl 367

1007 AGGTGGGAGCATGTGCATGATGATTTGGGATCCAAATATTCGTT.... 1051

367 YelYrYsArpArTrIleYrYrAspLeuGlyIleYrYsYrSerPheTrH 384

1052TAC 1054

384 IeGlnUleuArgAspTrHrclYrYrYrPheUleuLeuProGlnUArgYr 400

1055 ATCAAAACCCACCTGTAGAGAAGCTTTTGCCTGCTCTCTAAATAGSTTG 1104

401 TrleYsProThrCysArgGlnUalArPheAlaValaSerYsIleAlaTr 417

1105 GCATGTCATAGGAATGTT 1123

|||||

```

417 PHISVal11leArgAsnVal 423
seq_name: pir2:A32129
seq_documentation_block:
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
C:Accession: A32129, S17543
R:Clausen, E.; Gaidell, S.J.; Craik, C.S.; Macdonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A:Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comp
A:Reference number: A92693; MUID:89034324
A:Accession: A32129
A:molecule type: DNA
A:Residues: 1-415 <CLAF>
A:Cross-references: GB:M23599; GB:J04041; MID:g203293; PIDN:AAA40872.1; PID:g203295
R:Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A:Title: Distribution of manganese in rat pancreas and identification of its primary
A:Reference number: S17543; MUID:91378950
A:Accession: S17543
A:molecule type: protein
A:Residues: 109-116, X, 118-130 <KOD>
C:Genetics:
A:Gene: CPB
A:introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase
F:354,376/active site: Tyr, Glu #status predicted

alignment_scores:
Quality: 695.50 Length: 419
Ratio: 2.644 Gaps: 8
Percent Similarity: 62.768 Percent Identity: 37.709

alignment_block:
US-09-980-881-1 x A32129 ..
Align seg 1/1 to: A32129 from: 1 to: 415

18 ATGAAGCTTTGCAGCCTTGCAAGCTCTGTGAACCATGTTCTCTCTGTGA 67
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 MetLeuLeuLeuLeuAlaLeuValSerValAlaLeuAlaHisAlaSerGlu 17
68 GCAGCATGCTCTCGCGTCCAGAGTGGCCAAAGTTGTACCTGCTCTTCCCTA 117
||: ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
17 uGluHis.....PheAspGlyAsnAlaGlyAlaGlyAlaGlyAlaHis 31
118 GAACCTCTAGGCAAGTTCACAGATCTTACAGATCTTACTACCAACATATGAG 167
::: :::: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
31 ISGlyGluAspHisValAsnLeuIleGlnGluLeuAlaAsnThrLeuSerGlu 47
168 ATTGTCTCTGGCAGCCGCTAAACAGCTGACCTTATGTTGGAAGAAAACA 217
||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
48 IleAspPheThrPlysProAspSerAlaThrGlnValLysProLeuThrTh 64
218 AGTCCATTTTTGTAAATGCAATGATGTGCACAAATGTGAAGCCCAT 267
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 IValAspPheHisValLysAlaGluAspValAlaAspAlaGluAsnPhel 81
268 TAAATGTGACGGGAATTCATGCAGTGCTGTGCGGAGAGCTGGAGAT 317
||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
81 euGluGluAsnGluValHisThrGluValLeuIleSerAsnValAlaArgAsn 97
318 CTATTTCAACAGCAGATTTCCACAGACAGCTGACGCCCGAGGCTCGCG 367
::: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
98 AlaLeuGluSerGlnPheAspSerHisThr.....ArgAlaSerGlu 111
368 ATTCGACATGACAGATATCTACCTACCTAAATGAATGAAATGATCTTGATAG 417
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 yHisSerThrThrLysTyrAsnLysTrpGluThrIleGluAlaThrPleG 128

```



```

166 gLysAlaIlePheMetAspCysGlyIleHisAlaArgLutPrIleSerp 183
580 CTGCTTCTGCTTGTGCTGCTATA.....GGCCAT 608
|||||.....|
183 rolaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199
609 AAT.....|
200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPhetYrValLeuPr 216
612 .....CGAATGT 618
216 oValPheAsnValAspGlyTyrIleTrpSerTrpTrpGlnAspArgMet 233
619 GGAGAAAGAACCGTTCTTCTATGCGAACATCATGTCATGGAAACAGAC 668
|||||.....|
233 rPrArgLysAsnArgSerArgAsnGlnAsnSerThrCysIleGlyThrAsp 249
669 CTGAAATAGCACTTGTCTCCAAACACTGGTGAGGAGAGCTGATCCAG 718
|||||.....|
250 LeuAsnArgAsnPha...AspValSerTrpAspSerSerProAsnThrAs 265
719 TTCCTCATGCTCGAAACCTACTGTGAGCTTTATCCGTAGTCAGAACAG 768
::: ||| ::::| ||| |||
265 nysProCysLeuAsnValTyrArgGlyProAlaProGlnSerGluLysG 282
769 AAGTGAAGCAGTGGCTAGTCTTGTGAGAAAGAAATATCAACGATTTAA 818
|||||.....|
282 LuthrLysAlaValThrAsnPhelIleArgSerHisLeuAsnSerIleLys 298
819 GCATATACATGACATGATCATCTATCCCAACATATAGTGTTCATATTC 868
|||||.....|
299 AlatYrIleThrPheHisSerTyrSerGlnMetLeuIleProTyrGrl 315
869 CTATACACAGAGTAAAGCAAGAACCATGAGAACTGCTCTAGTAGCCA 918
|||||.....|
315 yTyrThrPheLysLeuProProAsnHisGlnAspLeuLysValAlaAla 332
919 GTGAGACATTCGCTGATTTGACAAACATAGTAAATACAGGATTTACA 968
|||||.....|
332 rGleLeuAlaThrAspAlaLeu...SerThrArgTyrGluThrArgTyrIle 347
969 CATGCCATGCGCTGACGAACCTTATCTAGCTGCTGAGGTGGAGACGA 1018
::: ||| ::::| ||| |||
348 TyrGlyProIleAlaSerThrIleTyrLysThrSerGlySerLeuAs 364
1019 TTGGATCTATGATTTGGCATCAAAATATTCGTTTAC.....ATCAAA 1054
|||||.....|
364 pTrpValTyrAspIleuGlyIleLysHisThrPheAlaPheGluLeuArg 381
1055 .....ATCAAA 1066
381 sPLysGlyLysSerGlyPheLeuLeuProGlnSerArgIleLysProThr 397
1067 TGTAGAGAGCTTTTGGCCCTGCTCTCTAAATAGCTTGGCATGCTTAC 1116
|||||.....|
398 CysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeu 414
1117 GAAT 1120
::: |||
414 sAsn 415

seq_name: plr2:A43929
seq_documentation_block:
carboxypeptidase A (EC 3.4.17.1) CPA3 precursor - human
N:Alternate names: mast cell carboxypeptidase A3
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence, revision 31-Dec-1993 #text, change 22-Jun-1999
C:Accession: A43929, A39246, A45759
R:Reynolds, D.S.; Gurley, D.S.; Austen, K.F.
J. Clin. Invest. 89, 273-282, 1992
A:Title: Cloning and characterization of the novel gene for mast cell carboxypeptidase A

```

```

A:Reference number: A43929; MUID:92105393
A:Accession: A43929
A:Molecule type: DNA
A:Residues: 1-417 <REV>
A:Cross-references: GB:M73716
A:Experimental source: mast cell
A:Note: The authors translated the codon GCG for residue 211 as Thr
R:Reynolds, D.S.; Gurley, D.S.; Stevens, R.L.; Sugarbaker, D.J.; Austen, K.F.; Serafi
Proc. Natl. Acad. Sci. U.S.A. 86, 9480-9484, 1989
A:Title: Cloning of cDNAs that encode human mast cell carboxypeptidase A, and compari
A:Reference number: A39246; MUID:90083291
A:Accession: A39246
A:Molecule type: mRNA
A:Residues: 1-417 <REV>
A:Cross-references: GB:M7717; NID:G179933; PID:AAA35652.1; PID:G179934
R:Goldstein, S.M.; Kaempfer, C.E.; Kealey, J.T.; Wintroub, B.O.
J. Clin. Invest. 83, 1630-1636, 1989
A:Title: Human mast cell carboxypeptidase. Purification and characterization.
A:Reference number: A45759; MUID:89214692
A:Accession: A45759
A:Molecule type: protein
A:Residues: 110-137 <GOL>
C:Genetics:
A:Gene: GDB:CPA3
A:Cross-references: GDB:125231; OMIM:114851
A:Map position: 3q21.3-3q25
A:Introns: 23/2; 48/3; 90/2; 124/3; 158/3; 192/3; 229/3; 260/1; 327/3; 356/1
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; z
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-109/Domain: activation peptide #status predicted <ACT>
F:110-417/Product: carboxypeptidase A, mast cell #status predicted <MAT>
F:176-179,304/Binding site: zinc (His, Glu, His) #status predicted
F:245-268/Disulfide bonds: #status predicted
F:356,378/Active site: Tyr, Glu #status predicted

alignment_scores:
Quality: 647.50 Length: 427
Ratio: 2.398 Gaps: 10
Percent Similarity: 63.232 Percent Identity: 33.488

alignment_block:
US-09-980-881-1 x A43929 ..

Align seg 1/1 to: A43929 from: 1 to: 417

6 ATTGCGTGGGATGAGAGCTTGGACGCTTGGACGCTTGTACCAATTGT 55
::: |||||::: ::|||:::
5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle...AlaProVal.. 19
20 .....ArpPheAspArgGluLysValAlaPhe 28
56 TCTCTTCTGTGAGACGATGCTTGCCTTCCAGCGTTCAGAGTGGCAAGTTTAC 105
|||||.....|
106 CTGCTCTTCTCTGAACCTCTGACGCAAGTTCAAGTTCTACAGAAATCTTACT 155
::: |||||::: ::|||:::
28 rGValLysProGlnAspGluLysGlnAlaAspIleIleLysAspLeuAla 44
156 ACAACATATGAGATTTGTTCTGTGGCAGCGCGTAAACAGCTGACCTTATGT 205
||| |||||::: ||| |||
45 LysThrAsnGluLeuAspPheThrPyrTrpGlyAlaThrHisHisValAla 61
206 GAAGAAAAAAGCAAGTCCATTTTGTAAATGACATCTGTATGTCAGCAATG 255
::: ||||| ||| |||||::: ::|||:::
61 AlaLysMetMetValAspPheArgValSerGluLysGlnAlaIle 78
256 TGAAGGCCCATTTAAATGATGAGCGGAATTCATGCAAGTGTCTGTGGCA 305
::: |||||::: ::|||:::
78 LeuLysThrAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuHis 94
306 GACGTGGAAGATCTTATTCACAGCAGATTTCC...AACGACACAGTCAG 352
|||||.....| |||||::: |||||:::

```


[illegible]

```

910 TGTGTGACCGATTAACAGAGTTCGTGCTATTGTGACAAAACAGTAAAAATACC 959
    :::::  :::::  :::  :::
328 lAeuLalAyslaIaThrVAlAysGIuLeu..AlASeLueHIselIyThr 343
    lAeuLalAyslaIaThrVAlAysGIuLeu..AlASeLueHIselIyThr 343
960 AGATATATACATGAGCGCATGCGTCACAGAAACCTATACAGTACCTCGAGAG 1009
    :::::  :::::  :::::  :::::  ||  ||||
344 LysTyrThrTyGclProGlylaIaThrThrlIeTyPrroAlaAlaGlyGI 360
    :::::  :::::  :::::  :::::  :::::  :::::
1010 TGGGACGATTTGGATCTATGATTGGGCATCAAAATTCGTTTTATATC.. 1057
    :::::  :::::  :::::  :::::  :::::  :::::
360 ySeArSpaSpTPrAlaIATyAspGlnGlyLeuArgTySeArPhetrrPhes 377
    :::::  :::::  :::::  :::::  :::::  :::::
1057 ..... 1057
377 lUeUaArgSpThrGlyArgTyGlyPheLeuLeuProGluSerGlnIle 393
    lUeUaArgSpThrGlyArgTyGlyPheLeuLeuProGluSerGlnIle 393
1058 AAACCCACCTGTAGAGAAAGCTTTTGGCCGTGTCCTAAATATAGCTTGCA 1107
    ::  |||||  :::::  :::::  |||||  :::
394 ArgAlaIaThrCyAspGlnGlnThrPheLeuAlaIleTySyrValAlaSerTy 410
    ArgAlaIaThrCyAspGlnGlnThrPheLeuAlaIleTySyrValAlaSerTy 410
1108 TGCATTAGGATGTT 1123
    TGCATTAGGATGTT 1123
410 rVAlleuGlnHIsIeU 415
    rVAlleuGlnHIsIeU 415

```

```

seq_name: pIrl:CPBOB
seq_documentation_block:
carboxypeptidase B (EC 3.4.17.2) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Apr-1994 #sequence_revision 24-Apr-1994 #text_change 02-Aug-1994
C:Accession: A93797; A92150; A00912
R:Titani, K.; Ericsson, L.H.; Walsph, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 72, 1666-1670, 1975
A:Title: Amino-acid sequence of bovine carboxypeptidase B.
A:Reference number: A93797; MUID:75217824
A:Accession: A93797
A:Molecule type: protein
A:Residues: 1-306 <TIP>
R:Schmidt, J.J.; Hirs, C.H.W.
J. Biol. Chem. 249, 3756-3764, 1974
A:Title: Primary structure of bovine carboxypeptidase B. Inferences from the location
A:Reference number: A92150; MUID:74260705
A:Accession: A92150
A:Molecule type: protein
A:Residues: 31-93;131-181;263-265;292-306 <SCH>
R:Schmid, M.F.; Herriott, J.R.
J. Mol. Biol. 103, 175-190, 1976
A:Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
A:Reference number: A92839; MUID:76265065
A:Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
R:Plummer Jr., T.H.
J. Biol. Chem. 244, 5246-5253, 1969
A:Title: Isolation and sequence of peptides at the active center of bovine carboxypep
A:Reference number: A92051; MUID:70007159
A:Contents: annotation; active site
R:Kimmel, M.T.; Plummer Jr., T.H.
J. Biol. Chem. 247, 7864-7869, 1972
A:Title: Identification of a glutamic acid at the active center of bovine carboxypept
A:Reference number: A92119; MUID:75061487
A:Contents: annotation; active site
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; z
F:63-76,135-158,149-163/Disulfide bonds: #status experimental
F:66,69,194/Binding site: zinc (His, Glu, His) #status experimental
F:246,268/Active site: Tyr, Glu #status experimental

```

```
alignment_scores:
  Quality: 589.50
  Ratio: 3.039
  Length: 3044
  Gaps: 6
Percent Similarity: 63.816
Percent Identity: 39.803
```


alignment_block:
US-09-980-881-1 x CPBOB ..

Align seg 1/1 to: CPBOB from: 1 to: 306

```

375 TATGACAGATATCAGTCACCTAAATGAATCTATCTTGATAGATTTAT 424
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6 TyrlulysylrnsnspntrpGluThrIleGluIaIaIaIaIaIaIa 22
425 AACTGAGAGCATCTGATGATGCTTACAAAATCCACATGATCTCAT 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 IAlaSerGIuAsnProAspleuIleSerAlaIleGlyThrThr 39
475 TTGAGAAGTACCCACTCTATGTTTAAAGTTTCTGGAAGAAACAACA 524
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39 heIeuGIaSnThrIleTyrLeuIeuIaVal...GlyLysProGIySer 54
525 GCCAAAATGCCATATGATGATGATGATGATGATGATGATGATGAT 574
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 AsnLysProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 71
575 CTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 eSerProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 88
602 ..... 602
88 LyaGIaGIuIleHISmetThrGIuPheIeuAspLysLeuAspPheTyrVal 104
603 .....GCCCAT.....AATCG 613
105 LeuProValaIaSnIleAspGIyTyrIleTyrThrTrpThrThrAsnAr 121
614 AATGTGAGAAAGAACCGCTCTTCTGATGCGAACAATCATGTCAGGAA 663
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 gMetTrpArgLysThrArgSerThrArgAlaGIySerSerCysThrIyT 138
664 CAGACCTGAATAGCAACTTTGTCTCCAAACACTGTGTGAGGAAGTGA 713
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 hTrsAspLeuAsnArgAsnPhe...AspAlaGIyTrpCysSerIleGIyAla 153
714 TTCAGTTCCCATGCTCGGAAACCTACTGTGACTTTATCTGATGTCAGA 763
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 SerAsnAsnProCysSerGIuThrTyrCysGIySerAlaIaIaGIuSerGI 170
764 ACCGAGAGTGAAGGACAGTGGCTAGTTCTTGAGAGAATAATATCAACA 813
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 uLysGIuSerLysAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 187
814 TTAAAGCAATACATCAGCATGATTCATATCCACGATATAGTGTTCGA 863
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 IeLysAlaIaTyrLeuThrIleHISserLysSerGIuMetLeuTyrPro 203
864 TATTTCTATACAGAACTAAAGCAACATGAGAGAACTGTCTAGT 913
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 TyrSerTyrAspTyrTrpLysLeuProLysAsnAsnValGIuIeuAsnThrIle 220
914 AGCGAGTGAAGCATGCTGCTATGATGACAAAACCTAGTAAATATCAAGT 963
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 uAlaIaLysGIyAlaIaIaLysLysLeu...AlaSerLeuHISGIyThrThrT 236
964 ATACACATGGCATGGCTCAGAAACCTTATATACCTGCTGAGAGTGGG 1013
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 ySerTyrGIyProGIyAlaIaThrThrIleTyrProAlaSerGIyGIySer 252
1014 GAGGATGGATCTATGATGTTGGGCAACAATATATCTGTTACATC..... 1057
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 AspAspTrpAlaTyrAspGIuGIyIleLysTyrSerPheThrPheGIuIe 269
1058 .....AAC 1061
269 uArgAspLysGIyArgTyrGIyPheValLeuProGIuSerGIuIleIaIa 286

```

```

1062 CCACCTGTAGAGAAAGCTTTGCCCGCTGTCTCTAAATAGCTTGCAATGTC 1111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 roThrCysGIuGIuThrMetLeuAlaIleLysTyrValThrSerTyrVal 302
1112 ATTAGCAATGTT 1123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 LeuGIuHISLeu 306
seq_name: pIrf2:A38395

```

```

seq_documentation_block:
mast cell carboxypeptidase (EC 3.4.-.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C:Accession: A38395; A33118
R:Collection: K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A:Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme a
A:Reference number: A38395; MOID:91105153
A:Accession: A38395
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-309 <COT>
R:Collection: K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A:Reference number: A33118
A:Accession: A33118
A:Status: preliminary
A:Molecule type: protein
A:Residues: 2-309 <COT>
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase
F:248,270/Active site: Tyr, Glu #status predicted

```

alignment_scores:
Quality: 567.50 Length: 303
Ratio: 2.925 Gaps: 6
Percent Similarity: 64.026 Percent Identity: 39.604

alignment_block:
US-09-980-881-1 x A38395 ..

Align seg 1/1 to: A38395 from: 1 to: 309

```

375 TATGACAGATATCAGTCACCTAAATGAATCTATCTTGATAGATTTAT 424
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8 TyrlalysylrnsnspntrpAsnLysIleValaIaIaIaIaIaIaIaIa 24
425 AACTGAGAGCATCTGATGATGCTTACAAAATCCACATGATGATCTCAT 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24 tValGIuLysHISProGIuMetValSerArgIleLysIleGIySerThrY 41
475 TTGAGAAGTACCCACTCTATGTTTAAAGTTTCTGGAAGAAACAACA 524
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41 alGIaAspAsnProIeuTyrValIeuLysIle...GIyArgLysAspGIy 56
525 GCCAAAATGCCATATGATGATGATGATGATGATGATGATGATGATGAT 574
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 GIuArgLysAlaIlePheMetAspCysGIyIleHISAlaIaIaIaIaIaIa 73
575 CTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 IserProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 90
604 GCCATTAAT..... 611
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 LysAsnAsnIleMetThrLysLeuIeuAspArgMetAsnPheTyrVal 106
612 .....CG 613
107 LeuProValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 123

```

```

614 AATGTGAGAAAGAACCGTTCTTTATGCGAAACATCATGTCATGCGAA 663
|||||
123 gmetrparlylsasnarlyserlysasnpserSerThrCys116GlyT 140
|||||
664 CAGACCTGAATACCACTTTGCTCCAAACACGTGCTGAGAAAGGTCA 713
|||||
140 hrAspleuasnarganphe...AspValSerThrAspSerSerProasn 155
|||||
714 TCCAGTTCTCATGCTCGGAAACCTGAGACTTATTCCTGAGTCAGA 763
|||||
156 ThrAspAsnProCysLeuSerValTyrArgGlyProAlaProGluSerG 172
|||||
764 ACCAGAGGAAAGGACGAGTGGCTGAGTTCTTGAGAAAGAAATATACACAGA 813
|||||
172 uylsGlyThrLysAlaValThrAsnPhelIeaArgSerHisLeuAsnSerI 189
|||||
814 TTTAAACATACATCAGCATGATTCATCTCCAGCATATAGTGTTCGA 863
|||||
189 lelyalalalyrIleThrPhelIeaSerTyrSerGlnMetLeuLeuPhePro 205
|||||
864 TATTCCTATACACAGAACTAAAGCAAAACCATGAGAACTGCTCTAGT 913
|||||
206 TylGlyTyrThrIleLysLeuProProAsnHisGlnAspLeuLeuLysVa 222
|||||
914 AGCCAGTGAAGCAGTTCGTGCTATTGACAAACTACTATAAAATACACAGT 963
|||||
222 lAlaArgGlyIleAlaThrAspValLeu...SerSerArgTyrGluThrArgT 238
|||||
964 ATACACATGCGCATGCGTCAGAAACCTTATACCTGCTGAGAGGTGGG 1013
|||||
238 yrlIeTyrGlyProIleAlaSerThrIleTyrLysThrSerGlySerSer 254
|||||
1014 GACGATGGATCTGATGTTGGGATCAAAATATTCGTTTAC..... 1054
|||||
255 LeuAspTrpAlaTyrAspLeuGlyIleLysHisThrPhelAlaPheGlu 271
|||||
1055 .....ATCAAC 1061
|||||
271 uArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysP 288
|||||
1062 CCACCTGTGAGAGAGCTTTGGCGCTGCTTAAATACCTTGCGCAATGC 1111
|||||
288 roThCysLysGlyLysThrMetLeuSerValLysPheIleAlaLysTyrIle 304
|||||
1112 ATTTAGGAAT 1120
|||||
305 LeuLysHis 307
|||||
seq_name: pLr2:A32128
seq_documentation_block:
carboxypeptidase A2 (EC 3.4.17.15) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999
C:Accession: A32128
R:Gardell, S.J.; Craik, C.S.; Clausen, E.; Goldsmith, E.J.; Stewart, C.B.; Graf, M.; Rut
J. Biol. Chem. 263, 17828-17836, 1988
A:Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and ex
A:Reference number: A32128; M01D:89034323
A:Accession: A32128
A:Molecule type: mRNA
A:Residues: 1-417 <GAR>
A:Cross-references: GB:M23719; GB:M23721; GB:J04043; NID:g341029; PIDN:AAA40956.1; PID:g
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase
F:356/378/Active site: Tyr, Glu #status predicted

```

```

alignment_scores:
  Quality: 553.00      Length: 409
  Ratio: 2.194        Gaps: 13
Percent Similarity: 61.614      Percent Identity: 33.496

```

```

alignment_block:
US-09-980-881-1 x A32128 ..
Align seg 1/1 to: A32128 from: 1 to: 417
18 ATGACGTTTGCAGCCTTGCGCTGCTTACCCATGTTGCTCTGTGA 67
|||||
1 MetArgLeuThrLeuLeuAlaLeuLeuGlyTyrIleTyrCysG 17
68 GCAGCATGCTGCGGTCCAGAGTGGCCAGTTCTTACGCTGCTCCGA 117
|||||
17 nGlu.....ThrPheValGlyAspGlnValLeuGluIleLeuPro 31
118 GAACCTCTAGGCAAGTTCAGATCTCAGAAATCTTACTACACA..... 161
|||||
31 ernHisGlnGlnGlnIleArgThrLeuLeuGlnLeuGlnAlaGlnLys 47
162 TATGACATGTTCTCTGCGACCGGTACAGCTGACCTTATGTTGAAGA 211
|||||
48 LeuGlnLeuAspPheTyrLysSerProThr.....IleProGlu 60
212 AAACAGATCCCATTTTGTGTAATGATCATGATGTCGACAAATGCAAG 261
|||||
60 yGluThrValHisValAlaGlyValProPheAlaSerIleGlnAlaValLysV 77
262 CCCATTAAATGTGACGCGGAATTCATGCAAGTCTTGTGCGACGACGT 311
|||||
77 AlPheLeuGlnSerGlnGlyIleAspTyrSerIleMetIleGluAspVal 93
312 GAAGATCTTATTCACAGCAGATTTCCAAC.....GACACAGTCAG 352
|||||
94 GlnValLeuLeuAspGlnGlnIleArgGlnGlnMetLeuPheAsnGlnGln 110
353 CCCCCGAGCCTCCGATCTGATGATGACAGATGATGCTGCTGCTGCTGAA 402
|||||
110 gGluArgGlyGlyAsnPheAsnPhelValAlaTyrHisThrLeuGluLys 127
127 letyGlnGlnMetAspAsnLeuValAlaGlnAsnProGlyLeuValSer 143
403 TCTATTCCTTGATAGATTTATTAAGTGAAGCATCCATGATGCTTACA 452
|||||
143 letyGlnGlnMetAspAsnLeuValAlaGlnAsnProGlyLeuValSer 143
453 AAATTCACATTCGATCCATGATTTGAGAAATGACCCATGATGTTTAA 502
|||||
144 LysValAsnLeuGlySerSerPheGlnAsnArgProMetAsnValLeuLys 160
503 GGTTCCT...GGAAGAAGACAAACACCAAAATGATGATGATGATGACT 549
|||||
160 sPheSerThrGlyGlyAsp.....LysProAlaIleThrPheAsp 174
550 GTGGAATCCATGCGCAGAGATGATCTCTGCTTCTGCTGTGG... 596
|||||
174 lAgIylIleHisAlaArgGlnTrpValThrGlnAlaThrAlaLeuTrpThr 190
596 ..... 596
191 AlaAsnLysIleAlaSerAspTyrGlyThrAspProAlaIleThrSerLe 207
597 .....TTCATA.....GGCC 606
207 uLeuAsnThrLeuAspIlePheLeuLeuProValThrAsnProAspGlyT 224
607 AT.....AATCGAATGTGAGAAAGACGTTCTTTC 638
|||||
224 yValAlaPheSerGlnThrThrAsnArgMetTrpArgLysThrArgSerLys 240
639 TATGGAACAATCATTCATGCGAAGACACGATGATGACACTTGTCTC 688
|||||
241 ArgSerGlySerGlyCysValGlyValAspProAsnArgAsnTrp...As 256
689 CAACACATGCTGTGAGAGATGTCATCCATGCTGATGCTGCGAAGCT 738
|||||
256 pAlaAsnPhelGlyGlyProGlyAlaSerSerSerProCysSerAspSer 273

```



```

256 TGAAGCCCATTTAAATGTCAGCGGAATTCATGCACTGTCTGTCGCA 305
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    75 allysvalpheleuvalahisgllyleargtyrargllelellieg 91
306 GAGCTGAGAGATCTTAT.....CAACAGAGATTTCCAAAGCAC 346
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    92 AspyalginserleuLeuaspGluGluGluGluGluGluGluGlu 108
347 AGTCAGCCCCAGAGCTCCGATCGTAC...TATGAACAGTATCACTAC 393
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    108 nserArgalaaargSerThrAsnThrPheAsnTyraLaThrYHIsThrL 125
394 TAAATGAATCTATTCCTGATGAAATTTTAACTGAGAGCATCCGAT 443
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    125 euaspGluIleTyraSPheMetAspLeuValaGluHisProGln 141
444 ATGCTTACAAAATCCATGATGATCCTCATTTAGAAATGACCATCTA 493
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    142 LeuValSerTySleuGlnIleGlyArgSerTyGluGluYargProIleTy 158
494 TGTTTTAAAGTTTCTGGAAAAGAACACAGCCAAAATGCAATATGA 543
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    158 rValleuTySPheSer...ThrGlySerAsnArgProAlaIleTrpI 174
544 TTGACTGTGGAATCCATGCCAGAGAAATGATCTCTCTCTCTCTG 593
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    174 leaspGluGlyIleHisSerArgGluTrpIleThrGlnAlaThrGlyVal 190
594 TGGTTCATA..... 602
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    191 TrpPheAlaIleTySPheThrGluAspTyGlyGlnAspProSerPheThr 207
602 ..... 602
207 rAlaIleLeuaspSerMetAspIlePheLeuGluIleValThrAsnPro 224
603 .....GGCCATATCGAATGTGGAGAAAGAACCGT 632
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    224 spGlyPheAlaPheThrHisSerGlnAsnArgLeuTrpArgIleThrArg 240
633 TCTTTATCGAACAATATTCATCGAAGACAGCTGAAATGAACTT 682
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    241 ServaIThrSerSerSerLeuGlyValGlyValAspAlaAsnArgIle 257
683 TGTCTCCAAACACTGCTGAGAGAGCTATCCACTCTCTACTGCTGG 732
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    257 P...AspAlaGlyPheGlyLysAlaGlyAlaSerSerSerProCysSerG 273
733 AAACCTACTGTGAGCTTTATCTGAGTCAAGACCAAGCAAGCAAGT 782
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    273 luthrTyrlHisglTyrlaAlaAsnSerGluValGluValLysSerIle 289
783 GCTAGTTCTTGAGAAATAATCAACAGATTAAAGATATACATCAGAT 832
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    290 ValAspPheValLysAspHisGlyAsn...PheLysAlaPheLeuSerI 305
833 GCATTCATCTCCCAAGATATAGTGTTCATATTCCTATACAGAACTA 882
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    305 eHisSerTySerGlnLeuLeuLeuLeuTyProTyGlyTyThrThGlnS 322
883 AAAGCAAGACCATGAGAACTGTCTAGTAGCCAGTCAAGCAAGTCT 932
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    322 erIlePheProAspTyrlThrGluLeuAsnGlnValAlaLysSerAlaValGlu 338
933 GCTATTCACAAAATGTAATAATACACAGATTATACATGCGCATGGCT 982
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    339 AlaLeu...LysSerLeuTyGlyThrSerTyGlySerIleI 354
983 AGAAACCTTATACCTAGCTCCGAGAGTGGAGACATTTGATCTATGATT 1032
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    354 eThrThrIleTyGlnAlaSerGlyLysIleAspTrpSerTyAsnG 371
1033 TGGGCATCAAAATATTCGTTTACATCAAA..... 1060

```

```

371 lnglyIleTySerPheThrPheGluLeuArgAspThrGlyArgTy 387
1061 .....CCCACTGTGAGAAAGCTTT 1080
388 GlyPheLeuLeuProAlaSerGlnIleIleProThrAlaGlnIleThr 404
1081 TGGCGCTGTCTCTAAATAGCTTGGCATGTCATTAGAATGTT 1123
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    404 pLeuGlyValLeuThrIleMetGlnHisThrLeuAsnAsnLeu 418
seq_name: p1r1:CPRTA

```

```

seq_documentation_block:
  carboxypeptidase A (EC 3.4.17.1) precursor - rat
  C:Species: Rattus norvegicus (Norway rat)
  C:Date: 13-Jun-1983 #sequence
  C:Accession: A00911; B32129
  R:Quinto, C.; Quiroga, W.; Swain, W.F.; Nikovits Jr., W.C.; Strandring, D.N.; Pictet,
  Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982
  A:Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization
  A:Reference number: A00911; MUID:82105986
  A:Accession: A00911
  A:Molecule type: mRNA
  A:Residues: 1-419 <QUT>
  A:Cross-references: GB:J00713; NID:G203364; PIDN:AAA40893.1; PID:G203365
  R:Clausner, E.; Gardell, S.J.; Craik, C.S.; Macdonald, R.J.; Rutter, W.J.
  J. Biol. Chem. 263, 17837-17845, 1988
  A:Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comp
  A:Reference number: A92693; MUID:85034324
  A:Accession: B32129
  A:Molecule type: DNA
  A:Residues: 1-260, 'F', 262, 'W', 264-346, 'K', 348-419 <CLA>
  A:Gene: CP1A
  A:Genetics:
  A:Introns: 22/2: 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1
  A>Note: the authors translated the codon GTC for residue 196 as Ile
  C:Superfamily: carboxypeptidase
  C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; z
  F:1-16/Domain: signal sequence #status predicted <SIG>
  F:17-110/Domain: activation peptide #status predicted <PRO>
  F:111-419/Product: carboxypeptidase A #status predicted <MAT>
  F:179,182,306/Binding site: zinc (His, Glu, His) #status predicted
  F:248-271/Disulfide bonds: #status predicted
  F:358,380/Active site: Tyr, Glu #status predicted

```

```

alignment_scores:
  quality: 529.00      length: 436
  ratio: 2.124        gaps: 13
  percent similarity: 57.110      percent identity: 32.339.

```

```

alignment_block:
  US-09-980-881-1 x CPRTA ..

```

```

Align seg 1/1 to: CPRTA from: 1 to: 419

```

```

21 AAGCTTTCAGAGCTTCAGAGCTTCATGATGCTCTCTGTGAGCA 70
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    3 ArgLeuLeuIleLeuSerLeuLeuGlnAlaValLysGlyAsnGlnAs 19
71 GCATGCTTCGCGCTTCAGAGTGGCCAAAGTTCAGCTGCTTCCAGAA 120
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    19 n.....PheValGlyHisGlnValLeuArgIleSerAlaIla 32
121 CCTTAGAGAGTTCAGAGTTCAGAGTTCATGATGCTCTCTGTGAGCA 170
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    32 spGlnAlaGlnValGlnValGlnValGlnValGlnValGlnValGln 48
171 GTTCTC.....TGGCAGCGGTAACAGCTTATGTTGAAAGAAA 214
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    49 GlnLeuAspPheThrArgAspAlaAlaArgAlaGlyIleProIleAsp 65

```

```

215 ACAAGTCATTTTGTAAATGCATGTCATGACAAATGGAAGCC 264
    :||| |||
    :||| |||
65 1atgvalprohe.....Proserileglnservallysl 78
265 ATTAAATGAGCGGAATTCATGCATGCTGTGGCAGGATGGA 314
    :||| |||
    :||| |||
78 heuenglutyrlngllylesertygluilemetilegluaspvalgl 94
315 GATCTTAT.....CAACAGCAGATTTCACAGCACAGCAGCC 355
    :||| |||
    :||| |||
95 leuleuenspuglulysglnglmetserlarheglinalaayal 111
356 CCGAGCTCCGCATGTCAC...TATGAACGATACCTACTAATGA 402
    :||| |||
    :||| |||
111 aleuserthrasperheasnlylathrlylnisthrleuaspugl 128
403 TCTATCTGGATGAATTTATLACTAGAGGCGATCCGATATGCT 452
    :||| |||
    :||| |||
128 letygluprhemetasprleuvalalaglunhlsprogluvalser 144
453 AAAATCCATTTGATCCTCATTTAGAAAGTACCACCTATGTTTAA 502
    :||| |||
    :||| |||
145 lylsleeglntlegluasnthrphesgluglyargprolelnisval 161
503 GCTTTCGGAAGAACACAGCCAAAATGCCATATGATGATGACGTG 552
    :||| |||
    :||| |||
161 spheer...Thrglyglthrlnarargprolailetrpilesarg 177
553 GAATCCATGCGAGAAATGATCTCTGCTGCTGCTGCTGCTGCT 602
    :||| |||
    :||| |||
177 lyllelnserserarglutrpralthrlnalaserglvaltrphea 193
603 GGCAT..... 608
194 lylsvalthrlyasnpyrglygluasprrhphethralvalle 210
608 ..... 608
210 uasprasmelaspriheleugluilevalthrlnaspraspelyphea 227
609 .....AATGCATGTGAGAAAGAACCGTCTTCTCAT 641
    :||| |||
    :||| |||
227 latytrhlnslslythrlnarhmettrprarglythrlnaserhlnsthr 243
642 GCGAACAATGTCATGCGAACAACGTCGATAGCAATTTGCTCCAA 691
    :||| |||
    :||| |||
244 glndlyserleucysvalglyvalasprrohsnargasn..... 256
692 ACACTG.....TGTGAGGAAGTGCATCCAGTTCCTCATGCT 729
    :||| |||
    :||| |||
257 .....TTPAspAlagllyleuglylalsalglalaserhsnprcys 272
730 CGAACAACCTACTGTGACTTATCTGAGTACGAGAACGAGTGAAGCA 779
    :||| |||
    :||| |||
272 ergluptrtyrargllylurphesrprohsnsergluvalgluvalys 288
780 GTGGCTAGTTTCTGAGAAAGATATCAACCGATTAAGCATACATG 829
    :||| |||
    :||| |||
289 lilevalasprhevalthrserhlnslyasn...lleylsalapheliese 304
830 CATGCATTCATACCCACGATATAGTGTTCATATTCCTATACAGCA 879
    :||| |||
    :||| |||
304 rlelnlsertyrsergluileuvalthrprotyrlylthrlyserg 321
880 GTAAAGCAAAAGACCATGAGGAGTGTCTAGTAGCAGAGGAGCAGT 929
    :||| |||
    :||| |||
321 luprrolarproaspelinalagluleuaspuglndeualalysseral 337
930 CGTGCATTTGCAAAACTAGTAAATACAGGATACATGCGCATGG 979
    :||| |||
    :||| |||
338 ThrAlaleu...Thrserylnhslgllythrlnuphelsyrglyser 353
980 CTCAGAAACCTTATACCTAGCTCTGAGAGGTGGGAGCATTTGAT 1029

```

```

353 eileasprhrleryrglnalaserglyserthrlnleasprthrlyrs 370
    :||| |||
    :||| |||
1030 ATTTGGCATCAAAATATTCGTTTACATCAAC..... 1061
    :||| |||
    :||| |||
370 erglndlylleuslyrserythrlnphesgluileuargAspThrlyeu 386
1062 .....CCACCTGTAGAGAGCTTTGGCCTGTCTCTAAATA 1099
    :||| |||
    :||| |||
387 ArgglyPheleuleuProlnaserclnleleleProthrlnalaglul.. 402
1100 GCTTGGCATGTTCATTTAGAAATGTTAATGCCCTGATTTATACATTTG 1149
    :||| |||
    :||| |||
403 .....ThrrpleualaleuThrlnlemetasprlnsthrvallysh 417
1150 TTCGCTAT 1157
    :||| |||
417 lsprrtyr 419
seq_name: p1r2:A56171
seq_documentation_block:
carboxypeptidase A2 (EC 3.4.17.15) precursor - human
N/Alternate names: pancreatic carboxypeptidase A2
C/Species: Homo sapiens (man)
C/Date: 28-Apr-1995 #sequence_revision 11-Aug-1995 #text_change 22-Jun-1999
C/Accession: A56171; S02809; S71395
R/Cataus: L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Pulgarer, A.; Billeter, M.
J. Biol. Chem. 270, 6651-6657, 1995
A/Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. CD
A/Reference number: A56171; M01D:9520457
A/Accession: A56171
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-417 <CAT>
A/Cross-references: GB:019977; NID:q790226; PID:AAA74425.1; PID:q790227
A/Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue
R/Pascual, R.; Burgo, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A/Title: Purification and properties of five different forms of human procarboxypepti
A/Reference number: S02809; M01D:89153096
A/Accession: S02809
A/Molecule type: protein
A/Residues: 'S',18-36,'N',38-43 <PAS>
R/Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Wal
Arch. Biochem. Biophys. 332, 8-18, 1996
A/Title: Expression and characterization of human pancreatic preprocarboxypeptidase A
A/Reference number: S71394; M01D:96400327
A/Accession: S71395
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-301,'T',303-338,'SR',341-417 <LAD>
C/Genetics:
A/Gen: GDB:CPA2
A/Cross-references: GDB:125230; OMIM:600688
A/Map position: 7q32-7qter
C/Superfamily: carboxypeptidase
C/Keywords: hydrolase; metallo-carboxypeptidase; zinc
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-110/Domain: activation peptide #status predicted <ACP>
F:111-417/Product: carboxypeptidase A2 #status predicted <MAT>
F:177-180,304/Binding site: zinc (His, Glu, His) #status predicted
F:246-269/Disulfide bonds: #status predicted
F:356,378/Active site: Tyr, Glu #status predicted
alignment_scores:
Quality: 508.00 Length: 394
Ratio: 2.153 Gaps: 11
Percent Similarity: 59.898 Percent Identity: 32.487
alignment_block:
US-09-980-881-1 x A56171 ..

```

Align seg 1/1 to: A56171 from: 1 to: 417

```

72 CATGCTTC.....GCCTCCAGAGTGGCCAAAGTTCTAGTGGCTCT 112
   |||||.....:|||||:
13 HisIeTyrcysLeuGluThrPheValGlyAspIleValLeuGluIleVal 29
   |||||.....:|||||:
113 TCCAGAACCTCTAGGCAAGTTCAAGTTCTACAGAACTTACTACAAACAT 162
   |||||.....:|||||:
29 IProSerAsnGluGlnIleIleTyAsnLeuGlnIleuGluIleValGln 46
   |||||.....:|||||:
163 ATGAGATTGTCTC.....TGGACGGCGGTAACAGCTGACCTATATGTG 206
   |||||.....:|||||:
46 LHisIleGlnLeuAspPheTyrPheSerProIle.....Thr 58
   |||||.....:|||||:
207 AAGAAAAACAAAGTCATTTTCTTAATGTCATGATGTCGAAATGT 256
   |||||.....:|||||:
59 ProGlyGluThrIleHisValArgValProPheValAsnValGlnAlaVal 75
   |||||.....:|||||:
257 GAAAGCCCATTTAAATGTGAGCGGAATTCATGCGATGCTGTCTGGCAG 306
   |||||.....:|||||:
75 LysValPheLeuGluIleSerGlnGlyIleAlaIleTySerIleMetIleGlu 92
   |||||.....:|||||:
307 ACGTGAAGATCTTATCAACAGCAGATTCCAAAGCAGACAGATGAGCCG 356
   |||||.....:|||||:
92 spValGlnValLeuLeuAspIleAsnGlnGluMetIlePheAsnArg 108
   |||||.....:|||||:
357 CGAGCTCCGCATCG.....TACTATGAACAGATGACTCATCTATAA 397
   |||||.....:|||||:
109 ArgArgGluIleGlySerGlyAsnPheAsnPheGlyAlaIleTyHisIleGlu 125
   |||||.....:|||||:
398 TGAATCTATCTTGGATAGATTAATTAAGTGAAGGATCCGATATGC 447
   |||||.....:|||||:
125 UGIIleSerGlnGluMetAspAsnLeuValAlaGlnHisProGlyLeu 142
   |||||.....:|||||:
448 TTCAAAATATCCACATGATGCTCTCATTTGAGAAGTACCACCTATAGT 497
   |||||.....:|||||:
142 AlserIleValAsnIleGlySerSerPheGluAsnArgProMetAsnVal 158
   |||||.....:|||||:
498 TTAAAGGTTTCT...GCAAAAGAACAAACGCCAAATATCCATATGAT 544
   |||||.....:|||||:
159 LeuLysPheSerThrGlyGlyAsp.....LysProAlaIleTyrIle 172
   |||||.....:|||||:
545 TGACGTGGAAATCCATGCGCAGAAATGATGATCTCTCGCTTCTGCTGT 594
   |||||.....:|||||:
172 UAspAlaGlyIleHisIleAlaArgGluIleValIleGlnAlaThrAlaLeu 189
   |||||.....:|||||:
595 GGTTCATAGGC.....GCAAAAGAACAAACGCCAAATATCCATATGAT 605
   |||||.....:|||||:
189 rPheAlaAsnLysIleValSerAspTyrGlyLysAspProSerIleThr 205
   |||||.....:|||||:
605 ..... 605
206 SerIleLeuAspAlaLeuAspIlePheLeuProValIleHisAsnProAs 222
   |||||.....:|||||:
606 .....CATATGCAATGTGGAGAAACAAACGCT 633
   |||||.....:|||||:
222 pGlyTyrValPheSerGlnThrLysAsnArgMetIlePheGlySerHis 239
   |||||.....:|||||:
634 CTTTCTATCGCAAAATCATTTGATGCGAAGACCTGAAATGCACTT 683
   |||||.....:|||||:
239 eTyValSerGlySerLeuGlyValAlaAspProAsnArgAsnTyr 255
   |||||.....:|||||:
684 GTCCTCAAAACACGCTGTCGAGAGGATTCAGTTCTCATGCTCGCA 733
   |||||.....:|||||:
256 ...AspAlaGlyPheGlyIleProGlyAlaSerSerAsnProCysSerAs 271
   |||||.....:|||||:
734 AACCTCTGCTGACTTTATCTGAGTCAGAACCAAGGAAAGGAGGAGTGG 783
   |||||.....:|||||:
271 pSerTyrHisGlyProSerAlaAsnSerGluValGluValLysSerIle 288
   |||||.....:|||||:
784 CTAGTTTCTTGAGAAATATCAACAGATTAAGCATATCATCAGCATG 833
   |||||.....:|||||:

```

```

288 AlaPheIleLysSerHis...GlyLysValLysAlaPheIleLeu 303
834 CATTCATATCTCCAGCATATAGTGTTCATATTCATATCCATACAGAACTAA 883
   |||||.....:|||||:
304 HisSerTyrSerGlnLeuMetCysPheProTyrGlyTyrLysCysTyrIle 320
   |||||.....:|||||:
884 AAGCAAAACCCATGAGAACTGTCTCTAGTACCCAGTGAAGCAGTTGCTG 933
   |||||.....:|||||:
320 sLeuAspPheAspGluLeuSerGluValAlaGlnIleLysAlaGlns 337
   |||||.....:|||||:
934 CTATGCAAAACACTAGTAAATACAGGTATACACATGCCCATGGCTCA 983
   |||||.....:|||||:
337 eTleu...ArgSerLeuHisGlyIleTySerTyLysValGlyProIleCys 352
   |||||.....:|||||:
984 GAAACCTTATACCTAGCTCTGAGAGTGGAGAGATGATGATGATGATTT 1033
   |||||.....:|||||:
353 SerValIleTyrGlnAlaSerGlyGlySerIleAspTyrTyrAsp 369
   |||||.....:|||||:
1034 GGGCATCAATATATCTGTTTACATCAAC..... 1061
   |||||.....:|||||:
369 rGlyIleLysTyrSerPheAlaPheGluLeuArgAspThrGlyArgTyrG 386
   |||||.....:|||||:
1062 .....CCACCTGTAGAGACGTTTGGCG 1085
   |||||.....:|||||:
386 LysPheLeuProAlaArgGlnIleLeuPro 396
   |||||.....:|||||:
seq_name: pIrl:s29127
seq_documentation_block:
carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human
N/Alternate names: pancreatic carboxypeptidase A1
C/Species: Homo sapiens (man)
C/Date: 25-Feb-1994 #sequence revision 19-Jan-1996 #text change 18-Jun-1999
C/Accession: S29127; A34205; S08253; S02810; S71394; S02811
R/Catatus, L.; Villagas, V.; Pascual, R.; Aviles, F.X.; Wicker-Pianguart, C.; Puigser
Biochem. J. 287, 299-303, 1992
A/Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A
A/Reference number: S29127; MUID:93038569
A/Accession: S29127
A/Molecule type: mRNA
A/Residues: 1-419 <CAT>
A/Cross-references: EMBL:X67318; NID:935329; PIDN:CAA47732.1; PID:935330
R/Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.
Am. J. Hum. Genet. 46, 795-800, 1990
A/Title: Human carboxypeptidase A identifies a BglII RFLP and maps to 7q31-qter.
A/Reference number: A34205; MUID:90196012
A/Accession: A34205
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 330-396 <STE>
A/Note: the authors translated the codon CTG for residue 391 as Val
R/Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A/Title: Further studies on the human pancreatic binary complexes involving procarbox
A/Reference number: S08253; MUID:90169111
A/Accession: S08253
A/Molecule type: protein
A/Residues: 17-43; 'xxx', 114-135 <MOU>
R/Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A/Title: Purification and properties of five different forms of human procarboxypepti
A/Reference number: S02809; MUID:89153096
A/Accession: S02810
A/Molecule type: protein
A/Residues: 17-42 <PAS>
R/laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Wal
Arch. Biochem. Biophys. 332, 8-18, 1996
A/Title: Expression and characterization of human pancreatic preprocarboxypeptidase A
A/Reference number: S71394; MUID:96400327
A/Accession: S71394
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-419 <IAB>

```


alignment_scores:

Quality: 437.50 Length: 424
Ratio: 1.878 Gaps: 16
Percent Similarity: 54.953 Percent Identity: 30.425

alignment_block:

US-09-980-881-1 x T33527 ..

Align seg 1/1 to: T33527 from: 1 to: 528

```

18 ATGAGGCTTGCAGC...CTGCAGCTCCTTACCACATGCTCTCTCTG 64
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MethrlyscysglnlleuAlaSerleuethrTrpSerPheTrpVa 17
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
65 TGACGACCATGCTCTCCGCTTCAGAGTGC..... 95
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 LpheaRgHlvalleuAlaThrThrAspGlyArgAlaPhearglyss 34
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96 ..... 104
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 eRSeRAlaThrThrProPAsnGlnLysThrSerSerPheLysleu 50
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
105 GCTGCTCTTCCTAGACCTTAGACAGTTCAGTTCATGCTAGACATCT.. 152
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ArgLLeAsnProGlnThrGlnGlySerValLysTrpLeuArgSerLeu 67
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
153 .....ACTACACATATAGATTTGCTCTGCAGCCGCTACAC... 191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 RGlAspRSeRProGlnThrGlnLeuAspPheTrpGlnProProthAsnI 84
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
192 .....GCTGACCTATGCTGAAAGAAAAACAAGTCATTTT 230
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 legLyalAlleValAspLeuThrValAlaProAlaAspAlaProAlaPhe 100
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
231 GTAATTCATCTGATGTGCACAAATGTGMAAGCCCATTTAAATGAGCGG 280
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 Val.....LysAspLeuGlnSerLysLys..... 108
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 AATTCCATGAGTGTCTGCTGGCAGCTGGAAGATCTTATTCAACAGC 330
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
109 .lleSerTyrlleValAlaValAsnAspLeuSerLysAlalleGlnAsnG 125
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
331 AGATTTCAGACAGACAGATGACGCCGCCGAGCTCCGAGTCACTAGAA 380
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
125 luvArgGlySerAspLysPheTyAsnProValAlaGlyPheAlaTyAsp 141
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
381 CAGTATCACTACTAAATGAAATCTATTCTTGATAGTAATTTAATCTGA 430
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
142 LysTlYAsnSerleuGlnGlnlleGlnThrGlnMetLysArgLeuLys 158
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
431 GAGGCATCCGATATGCTTACAAAATCCACATGATGCTCCTCATTTGAGA 480
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
158 sglutyRProthMetlleThrleuIleAspIleGlnGlnSerHleGln 175
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
481 AGTACCACACTATGTTTAAAGTTTCTGGAAGAAAGAAACAA..... 524
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175 snArgThrleuLeuValMetLyslleThrGlyLysArgAsnProLeuGly 191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
525 GCCAAAATGACCATATGATGATGATGATGATGATGATGATGATGATGAT 574
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
192 SerLyslleSerMetTrpIleAspAlaGlylleHleAlaArgGlyLys 208
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
575 CMTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 eAlaProAlaThrAlaMet...TyrIleAlaHleGlnLeuLeuGlyT 224
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
608 ..... 608
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
224 yRGlAsnAspAlaThrValAlaLysLeuMetAspHleAspPheTyR 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
609 .....AA 610
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

241 lleuProValMetAsnProAspGlyTyRGlLysTrpSerArgGlyLys 257
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
611 TCGAATGCGGAGAAAGAACGCTCT.....TTCT 639
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
257 nArgMetTrpArgLysAsnArgSerProAlaLysCysAlaArgGlnThr 274
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
640 ATCGACCAATCATTCATGATCGACAGACACCTGAATACCAATTTGCTCC 689
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
274 heserThrValCysCysSerGlyValAspLeuAsnArgAsnPhe...Asp 289
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
690 AAACACTGGTGTCAGAAAGTGCATCCAGTTCTCATGCTCGGAACCTA 739
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
290 TrpPheTrpAlaSerThrGlySerSerSerAspProCysHleAspThr 306
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
740 CTGTGACCTTATTCCTGATGACAGACAGAAAGTGAAGCGAGTGGT 789
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
306 HleSglSerAlaAlaPheSerGlnProGlnSerGlnAlaValArgAsp 323
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
790 TCTTGAGAGAAATATCACACGATTAAGCATATACATGATGATGATCA 839
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
323 heLeuGlnGlnAsnThrProGlu.....AlaPheIleSerleuHleSer 337
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
840 TACTCCAGCATATAGTGTTCATATCTCATATACAG...AGTAAAG 886
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
338 TySerGlnMetTrpLeuIleProTyGlnHleArgLysGlnSerTyR 354
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
887 CAAGAAC...CATGAGGAACTGCTCTAGTACGACAGTGAAGTTCGCTG 933
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 OGlnAspTyHleThrGlyLeuArgProLeuAlaValAlaThrLys 371
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
934 CTATTGACAAAACCTAGTAAATAATGACAGTATACATGCGCATGCTCA 983
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 lAsuTyRGlLeu...TyRGlTyThrLysTyRlValGlyThrGlyAla 386
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
984 GAAACCTTATCCATGCTCTGAGAGGAGGAGGAGCATGATCAT...CA 1030
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
387 AspLeuMetTyRGlLysAspGlyGlySerHleAspTrpAlaLysGly 403
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1031 TTTGGGCATCAAAATATTCGTTT 1052
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
403 nLeuLysValProTyRAlaTyR 410
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: p1r2:T33526

seq_documentation_block:

hypothetical protein T06A4.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T33526

R:Wu, X.; Tin-Mollam, A.; Ozersky, P.; Wilson, R.

submitted to the EMBL Data Library, October 1998

A:Description: The sequence of C. elegans cosmid T06A4.

A:Reference number: Z1364

A:Accession: T33526

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-667 <WUX>

A:Cross-references: EMBL:AF098994; PIDN:AA67474.1; GSPDB:GN00019; CESP:T06A4.3

C:Genetics:

A:Gene: CESP:T06A4.3

A:Map position: 1

A:Introns: 50/3; 96/2; 167/3; 236/3; 276/3; 315/1; 365/3; 385/3; 444/1; 487/1; 582/3

alignment_scores:

Quality: 400.50 Length: 483
Ratio: 1.655 Gaps: 17
Percent Similarity: 50.104 Percent Identity: 27.122

alignment_block:

US-09-980-881-1 x T33526 ..

OM of: US-09-980-881-1 to: SwissProt_40:* out_format : pfs
Date: Sep 18, 2002 4:43 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frim+asp_model -DEV=xlh
-O=/cgn2_1/USPto_spool/US09980881/runat_16092002_140042_8937/app-query.fasta_1.1655
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=Dissim62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881@cgn1_1.22 -NCPU=6 -ICPU=3 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-1
Query length: 1573
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 37.420000

score_list:

Sequence	Strid	Orig	Zscore	Escore	len	Documentation
Swissprot_40:CBPB_RAT	695.50	1078.73	5.2e-53	415	1	P19223 rattus norvegicus (rat)
Swissprot_40:CBPB_MOUSE	655.00	1015.12	1.8e-49	417	1	P15089 mus musculus (mouse)
Swissprot_40:CBPB_HUMAN	652.50	1011.20	3.0e-49	417	1	P15086 homo sapiens (human)
Swissprot_40:CBPC_HUMAN	647.50	1003.35	8.2e-49	417	1	P15088 homo sapiens (human)
Swissprot_40:CBPB_CANFA	640.50	992.38	3.4e-48	416	1	P55261 canis familiaris (dog)
Swissprot_40:CBPB_PIG	629.50	975.45	3.1e-47	401	1	P09955 sus scrofa (pig)
Swissprot_40:CBPB_BOVIN	589.50	915.11	9.2e-44	306	1	P00732 bos taurus (bovine)
Swissprot_40:CBPB_RAT	567.50	880.49	7.8e-42	309	1	P1961 rattus norvegicus (rat)
Swissprot_40:CBPB_RAT	553.00	855.02	1.5e-40	417	1	P19223 rattus norvegicus (rat)
Swissprot_40:CBPB_BOVIN	536.00	828.29	4.6e-39	419	1	P00730 bos taurus (bovine)
Swissprot_40:CBPB_RAT	530.00	818.88	1.5e-38	419	1	P00731 rattus norvegicus (rat)
Swissprot_40:CBPB_HUMAN	508.00	784.39	1.3e-36	417	1	P48052 homo sapiens (human)
Swissprot_40:CBPB_HUMAN	489.50	755.31	5.4e-35	419	1	P15085 homo sapiens (human)
Swissprot_40:CBPB_HUMAN	483.00	745.06	2.0e-34	421	1	O91442 homo sapiens (human)
Swissprot_40:CBPB_SIMV	409.00	631.85	5.6e-28	304	1	P42788 simulium vittatum (b)
Swissprot_40:CBPA_ANOGA	386.00	592.56	6.0e-26	433	1	O02350 anopheles gambiae (a)
Swissprot_40:CBPB_ASTFL	369.50	569.88	1.6e-24	303	1	P04069 asciscus fluvialis (f)
Swissprot_40:YH2_YEAST	299.50	456.85	2.2e-18	430	1	P38836 saccharomyces cerevisiae (c)
Swissprot_40:CBPT_THEWU	277.50	422.44	1.8e-16	424	1	P23068 thermocactinomycetes vulva (t)
Swissprot_40:CBPS_STRGR	277.00	421.10	2.1e-16	451	1	P18143 streptomyces griseus (g)
Swissprot_40:CBPS_STRCP	252.50	382.99	2.8e-14	434	1	P39041 streptomyces capreolatus (c)
Swissprot_40:ORCL_SCHPO	118.50	168.26	0.0159	707	1	P54789 schizosaccharomyces f (s)
Swissprot_40:ORPO_PVHR	112.00	148.81	0.0693	1967	1	P17965 potato virus m (strat)
Swissprot_40:PER2_HUMAN	108.50	147.37	0.1305	1255	1	O15055 homo sapiens (human)
Swissprot_40:YQCT_BACSU	100.00	144.92	0.5964	376	1	P54497 bacillus subtilis (hy)
Swissprot_40:TRSH_CHICK	99.50	142.62	0.6775	445	1	P70080 gallus gallus (chick)
Swissprot_40:TRSH_ANGAN	98.50	140.21	0.8408	488	1	O42091 anguilla anguilla (e)
Swissprot_40:CBGA_BACNP	98.50	132.39	0.9649	1160	1	O45706 bacillus thuringiensis (b)
Swissprot_40:ITALI_SCHPO	97.00	135.09	1.19	663	1	P79005 schizosaccharomyces f (s)
Swissprot_40:TRSH_HUMAN	95.00	135.57	1.68	444	1	P17752 homo sapiens (human)
Swissprot_40:YQPT1_CAEEL	94.50	134.05	1.88	482	1	P41879 caenorhabditis elegans (e)
Swissprot_40:YQ31_YEAST	94.50	130.12	2.01	744	1	P53202 saccharomyces cerevisiae (c)
Swissprot_40:PYR1_DICDI	94.50	120.39	2.39	2185	1	P20054 dictyostelium discoideum (d)
Swissprot_40:YHNE_ECOLI	94.00	134.19	2.04	435	1	P78067 escherichia coli. putrefaciens (p)
Swissprot_40:SRFL_BACSU	93.50	114.34	3.16	3587	1	P27206 bacillus subtilis (s)
Swissprot_40:PHS1_DICDI	93.00	126.53	2.78	853	1	O00766 dictyostelium discoideum (d)
Swissprot_40:PU93_BUCAL	92.00	129.35	3.15	525	1	P57143 b bifunctional purine nucleoside phosphorylase (b)
Swissprot_40:VP3_ROTGI	92.00	125.16	3.39	835	1	P15736 simian 11 rotavirus (s)
Swissprot_40:STHA_ECOLI	91.50	129.66	3.41	465	1	P27306 escherichia coli. serot
Swissprot_40:ENV_SEV3L	91.50	122.91	3.85	982	1	P27399 simian foamy virus (ty

SwissProt_40:Y103_HUMAN + 91.00 132.93 3.52	297	1	O15006	homo sapiens (huma
SwissProt_40:CYB_PLAFA + 91.00 130.80 3.65	376	1	O02768	plasmodium falcipar
SwissProt_40:TRSH_RABIT + 91.00 129.30 3.75	444	1	P17290	oryctolagus cunicu
SwissProt_40:TLR_MOUSE + 90.50 120.73 4.75	1050	1	P58681	mus musculus (mou
SwissProt_40:SPTE_YEAST + 90.00 117.03 5.53	1451	1	P23615	saccharomyces cer

seq_name: SwissProt_40:CBPB_RAT

seq_documentation_block:

ID CBPB_RAT STANDARD; PRT; 415 AA.
AC P19223;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
GN CPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clausen E., Gardel S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family.";
RL J. Biol. Chem. 263:17837-17845(1988).
CC -1 CARBOXYPEPTIDASE: Peptidyl-L-lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M23950; AAA40872.1; -
CC EMBL: M23947; AAA40872.1; JOINED.
CC EMBL: M23950; AAA40872.1; JOINED.
CC EMBL: M23953; AAA40872.1; JOINED.
CC EMBL: M23952; AAA40872.1; JOINED.
CC EMBL: M23954; AAA40872.1; JOINED.
CC PUR: A32129; A32129.
CC HSSP: P09955; INSA.
CC MEROPS: M14.003; -
CC InterPro: IPR003146; Propep_M14.
CC InterPro: IPR000834; Zn_carboxypept.
CC Pfam: PF02244; Propep_M14; 1.
CC Pfam: PF00246; Zn_carboxypept. 1.
CC PRINTS: PR00765; CARBOXYPEPTASE.
CC PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
CC PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
CC HydroLase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
CC FT SGNL 1 13
CC FT PROPEP 14 108
CC FT CHAIN 109 415
CC FT METAL 174 174
CC FT METAL 177 177
CC FT METAL 302 302
CC FT ACT SITE 376 376
CC FT DISULFID 171 184
CC FT DISULFID 243 266
CC FT DISULFID 257 271
CC SO SEQUENCE 415 AA; 47515 MW; 8EA06CCAD30BEF2 CRC64;

alignment_scores: 695.50

Length: 419

Ratio: 2.644 Gaps: 8
Percent Similarity: 62.768 Percent Identity: 37.709

Alignment_block:
US-09-980-881-1 x CBPB_RAT

Align seg 1/1 to: CBPB_RAT from: 1 to: 415

```

18  ATGAGCTTTGAGCGCTTGACGCTTGTACCCATGTTCTCTGTGGA 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1  MetLeuLeuLeuLeuValSerValAlaLeuAlaHisAlaSerGI 17
68  GCAGCATGCTTGGCGTCCAGAGTGCCCAAGTTCAAGTCTCTCTCTTA 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17  uGIuHis.....PheaspGIysnaArgValItyArgValSerValH 31
118 GAACCTTAGGCAAGTTCAAGTCTCTACAGAACTTACTACACATATGAG 167
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
31  IselGIuAspHisValAsnLeuIleGIuLeuAlaAsnThrLysGIu 47
168 ATTGTTCTGTGGCAGCGCGGTACAGCTGACCTTATTGTGAAGAAAAACA 217
   ||| :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
48  IleAspPheThrLysProAspSerAlaThrGIuValLysProLeuThr 64
218 AGTCCATTTTGTAAATGCATCTGATGCGACAAATGTGAAGCCCAT 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64  rValAspPheHisValIleGIuAspValAlaAspValGIuAspPhe 81
268 TAAATGTGAGCGGAATTCATGCGATGCTCTGTGGCAGCGTGAAGAT 317
   || :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
81  euGIuLysnGIuValHisTyrGIuValLeuIleSerAsnValAlaGAsn 97
318 CTTATTCAACAGCAGATTCCAAAGCAGACAGTACGCCCGGACCTCCGC 367
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
98  AlaLeuGIuSerGIuPheAspSerHisThr.....ArgAlaSerGI 111
368 ATCGTACTATGAACAGTATCAGTCACTCAATGAATCTATCTGTGATAG 417
   : ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
111 yHisSerTyrThrLysTyrAsnLysTyrGIuThrIleGIuAlaThrIleGI 128
418 AATTATTAACTGAGAGCATCTGATATGCTTCAAAAAATCCATTTGGA 467
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
128 IngInValAlaThrAspAsnProAspLeuValThrGIuSerValIleGIy 144
468 TCCTCATTTGAGAAAGTACCCTCTATGTTTAAAGTTTCTGGAAGA 517
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
145 ThrThrPheGIuGIyArgAsnMetTyrValLeuLysIle...GIySer 160
518 ACAAAACAGCCAAAATATCCATATGATGATGACTGTGAATCCATCCAG 567
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
160 rArgProAsnLysProAlaIlePheHisLeuAspCysGIyPheHisAlaArg 177
568 AATGATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
177 LutrPheSerProAlaPheCysGIuTyrPheValArgGIuAlaValArg 193
602 ..... 602
194 ThrTyrAsnGIuIngInLeuHisMetLysGIuLeuAspGIuLeuAsp 210
603 .....GGCCAT..... 608
210 eTyrValLeuProValValAsnIleAspGIyValValTyrThrThrThr 227
609 ..AATGATGTGAGAAAGAACCGTCTTCTATGCGAACAATCAATTCG 656
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
227 yAspArgMetThrPargLysThrArgSerThrMetAlaGIySerSerCys 243
657 ATCGAAGACCTGATATGCAACTTGTCTCCAAACACTGCTGTGAGGA 706
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
244 LeuGIyValAlaArgProAsnArgAsnPhe...AsnAlaGIyTyrCysGI 259
707 AGGTGATCCACTTCTCTATGCTCGAAACCTACTGTGACTTTATCTCG 756

```

```

259  IGIyAlaSerArgSerProCysSerGIuThrTyrCysGIyProAlaProG 276
757  AGTCAGAAACAGAGAGAGGAGGCTAGTTCTTGAAGAAATATC 806
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276  LuserGIuLysGIuThrLysAlaLeuAlaAspPheIleArgAsnLeu 292
807  AACCGATTTAAAGCATATCATCAGCATGATTCATCTCCAGATATAGT 856
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
293  SerThrIleLysAlaTyrLeuThrIleHisSerTyrSerGIuMetIle 309
857  GTTCCATATTTCTATACACAGGTAAGCAAGCAAGACATGAGGAAGT 906
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
309  uTyrProTyrSerTyrAspTyrLysLeuProGIuAsnTyrGIuGIuLeu 326
907  CTCTAGTAGCAGGAAGCAGTTCGTCTTGTGCAAAATACTATAAAT 956
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
326  snAlaLeuValLysGIyAlaIleLysGIuLeu...AlaThrLeuHisGIy 341
957  ACCAGGTATACATGCGCATGGCTGCGTCAAGAACTTATCTGCTCTGG 1006
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342  ThrLysTyrThrTyrGIyProGIyAlaThrThrIleTyrProAlaIleGI 358
1007  AGGTGGGAGCAGATTGCATCTATGATTTGGCATCAAAATATCTTACAT 1055
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358  yGIySerAspAspThrSerTyrAspGIuGIyIleLysTyrSerPheThr 375
1056  C..... 1057
375  heGIuLeuAlaArgAspThrIngIyPhePheGIyPheLeuLeuProGIuSerGI 391
1058  ...AAACCCACCTGTAGAGAGAGCTTTGCCGCTGTCTTAAATAGCTTG 1104
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
392  IleArgGIuThrCysGIuGIuThrMetLeuAlaValLysTyrIleAlaAs 408
1105  GCATGTC 1111
408  nTyrVal 410
seq_name: SwissProt_40:CBPC_MOUSE
seq_documentation_block:
ID  CBPC_MOUSE  STANDARD;  PRT;  417 AA.
AC  P15089;
DT  01-APR-1990 (Rel. 14, Created)
DI  01-APR-1990 (Rel. 14, Last sequence update)
DE  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE  (Carboxypeptidase A3).
GN  CPB3.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=90062123; PubMed=2584208;
RA  Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA  Serafin W.E.;
RT  "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RL  novel member of the carboxypeptidase gene family."
RJ  J. Biol. Chem. 264:20094-20099 (1989).
CC  -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O -> peptide + L-
CC  amino acid.
CC  -1- SUBCELLULAR LOCATION: Secretory granules.
CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC  ZINC CARBOXYPEPTIDASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial

```


AC p15086: 060834; 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase B precursor (EC 3.4.17.2) (pancreas-specific protein) (PASP)
 DE (PASP)
 GN CPB1 OR CPB OR PCPB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
 RC TISSUE=Pancreas;
 RX MEDLINE=92129345; PubMed=1370825;
 RA Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S., French C.K.;
 RT "Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of pancreas-specific protein as pancreatic procarboxypeptidase B.";
 RL J. Biol. Chem. 267:2575-2581(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=99182241; PubMed=9524066;
 RA Aloy P., Calsus L., Vallegas V., Reverter D., Vendrell J., Aviles F.X.;
 RT "Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B.";
 RL Biol. Chem. 379:149-155(1998).
 RN [3]
 RP SEQUENCE OF 16-43.
 RC TISSUE=Pancreas;
 RX MEDLINE=89153096; PubMed=2920728;
 RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
 RT "Purification and properties of five different forms of human procarboxypeptidases.";
 RL Eur. J. Biochem. 179:609-616(1989).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O = peptide + L-lysine(or L-arginine).
 CC -1- TISSUE SPECIFICITY: PANCREAS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE ZINC CARBOXYPEPTIDASE FAMILY
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M81057; AAA66973.1; -
 DR EMBL: AJ224866; CAA12163.1; -
 DR PIR: S02812; S02812.
 DR PIR: A42337; A42332.
 DR HSSP: P09955; INSA.
 DR MEROPS: M14.003; -
 DR MIM: 114852; -
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; zn_carboxpept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; zn_carboxpept. 1.
 DR PRINTS: PR00765; CRBOXYPTASA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; zymogen; signal.
 FT SIGNAL 1 15
 FT PROPEP 16 110 ACTIVATION PEPTIDE.
 FT CHAIN 111 417 CARBOXYPEPTIDASE B.
 FT METAL 176 176 ZINC (BY SIMILARITY).
 FT METAL 179 179 ZINC (BY SIMILARITY).
 FT METAL 304 304 ZINC (BY SIMILARITY).

FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 173 186 BY SIMILARITY.
 FT DISULFID 245 268 BY SIMILARITY.
 FT DISULFID 259 273 BY SIMILARITY.
 FT CONFLICT 16 16 H -> A (IN REF. 1; AA SEQUENCE).
 FT CONFLICT 17 17 H -> Q (IN REF. 1; AA SEQUENCE).
 FT CONFLICT 37 37 H -> Q (IN REF. 3).
 FT CONFLICT 208 208 N -> D (IN REF. 1).
 FT CONFLICT 245 245 MISSING (IN REF. 1).
 SQ SEQUENCE 417 AA; 47366 MW; BBICF212D830305E CRC64;
 Alignment scores:
 Quality: 653.50 Length: 422
 Ratio: 2.394 Gaps: 8
 Percent Similarity: 64.692 Percent Identity: 33.649
 alignment_block:
 US-09-980-881-1 x CBPB_HUMAN ..
 Align seg 1/1 to: CBPB_HUMAN from: 1 to: 417
 24 CTTTCAGCCCTTGACGCTTGTACCATGTTCTCTGTGACGACA 73
 1 MetLeuAlaLeuValLeuValThrValAlaLeuAlaSerAlaHisI 17
 74 TGTCTTCGGG...TTCCAGAGTGGCCCAAGTCTAGCTGCTCTCTAGAA 120
 17 sGIyGLyGLuHisPheGLuGLyGLuValPheArgValAlaValGluNA 34
 121 CCTTCAGCAGGTTCAAGTCTTACAGATCTTACATCAATCAATAGAT 170
 34 sPGLuSnhIstIleAsnIleLeuArgIleuAlaSerThrThrGlnIle 50
 171 GTTCTGGGACGGGTAAACAGTGAAGCTTATTTGTGAAGAAACAACT 220
 51 AspPheArgValPheAspSerValThrGlnIleLeuProHisSerThrVa 67
 221 CCATTTTGTGTAATGACATGTATGTCGACAAATGGAAGCCATTGA 270
 67 lAspPheArgValValValAlaGlnIleAspPheValThrValGluAsnValLeu 84
 271 ATGTGACGGAATTCATGACAGTGTCTTGTCGACAGCTGGAAGATCT 320
 84 ySGLuSnhGLuLeuGlnIleValValLeuIleSerAsnLeuArgAsnVal 100
 321 ATTCAACAGCATTTTCCAAAGACAGTGAAGCCCGAGCTCGCATC 370
 101 ValGluAlaGlnPheAspSerArgVal.....ArgAlaThrGln 114
 371 GTACTATGAACTATCATCTCACTAAATGAATCTATTCGTGATGAAT 420
 114 sSerTyrGluLysTyrAsnLysTyrPheLysThrIleGluAlaThrPheGln 131
 421 TTATTAAGTGAAGGACATCTGATATGCTTACAAATAATCCATGTGATCC 470
 131 lValAlaThrGluAsnProAlaLeuIleSerArgSerValIleGlnThr 147
 471 TCATTGAGAGTACCACCTCTATGTTTAAAGTTTTCGGAAGAAAGACA 520
 148 ThrPheGlnIleArgAlaIleTyrLeuLeuVal...GlyLysAlaG 163
 521 AAGAGCAAAATGCCATATGATGATGATGATGATGATGATGATGATGAT 570
 163 yGluAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 180
 571 GGAATCT 602
 180 rPheSerProAlaPheCysGlnIlePheValArgGluAlaValArgThr 196
 602 602
 197 TyrGluArgGluIleGlnValThrGluLeuAsnLysLeuAspPheTyr 213

```

603 .....GGCCAT.....A 609
213 rvaileuProValleuAAsnleasprgLyrlleYThrThrprThryss 230
610 ATCGAAGTGGAGAAAGACCGCTTCTTATGGGAAACATCATTCGTC 659
230 eraArgpnerPrpArgLysThrArgSerThrHstHrGlySerSerCyslle 246
660 GGAACAGACCTGATATACCACTTGTCTCCAAACACTGCTGTGAGAGG 709
247 GlyThrspPrasnArgasnPhc...AspAlaGlyTrpCysGluIleG1 262
710 TGCATTCAGTTCCTATGCTCGGAACCTACTGTGACCTTATCTGAGT 759
262 yAlaSerArgasnPrpcysAspGluThrTyCysGlyPrcaAlaIaG1us 279
760 CAGACACGAGTGAAGGACGCTGCTAGTCTTCTTGAGACAAATATACAC 809
279 ergLulysGluThrLysAlaLeuAlaAspPheIleArgAsnLysIleuSer 295
810 CAGATTAAGCATCATCATCATCATCATCATCATCATCATCATCATCAT 859
296 SerleLysAlaTyrlleuThrIleHstSerTySerGlnMetIleTy 312
860 TCCATATTCCTATACAGCAAGTAAAGCAAGACCATGAGCACTGCTC 909
312 rPrctYSerTyrlaTyrlLysLeuGlyLuAsnAsnAlaGluLeuAAsnA 329
910 TAGAGCCAGTGAAGCACTGCTGCTATGTGCAAAACTAGTAAATATCC 959
329 laeuAlaLysAlaThrValLysGluLeu...AlaSerLeuHstGlyThr 344
960 AGGATATACATGGCCATGGCTGCTGAGAACTTATACCTAGCTGCGAGG 1009
345 LysTyThrTyrlGlyProGlyAlaThrThrIleTyPrcaAlaIaGlyG1 361
1010 TGGGAGCAGTATGATCTAATTTGGGCGATCAAAATATTCGTTACATC. 1057
361 ySerAspAspTrpAlaTyrlAspGlnGlyIleArgTySerPheThrPheG 378
1057 ..... 1057
378 luleuArgAspThrGlyArgTyGlyPheLeuLeuProGluSerGlnIle 394
1058 AAACCCACCTGTAGAGAGCTTTGCGCGCTGCTCTAAATAGCTTGCA 1107
395 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrlValAlaIaSerTy 411
1108 TGCATTAGCAAGTGT 1123
411 rvaileuGluHstIleu 416
seq_name: SwissProt_40:CBPC_HUMAN
seq_documentation_block:
ID CBPC_HUMAN STANDARD: PRT; 417 AA.
AC P15088;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=90083291; PubMed=2594780;
RA Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,

```

```

RA Austen K.F., Serafin W.E.;
RT "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT comparison of the protein with mouse mast cell carboxypeptidase A and
RT rat pancreatic carboxypeptidases."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mast cells;
RX MEDLINE=92105393; PubMed=1729276;
RA Reynolds D.S., Gurley D.S., Austen K.F.;
RT "Cloning and characterization of the novel gene for mast cell
RT carboxypeptidase A."
RL J. Clin. Invest. 89:273-282(1992).
RN [3]
RP SEQUENCE OF 110-417 FROM N.A.
RX MEDLINE=92333165; PubMed=1629626;
RA Natsukaki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsukaki M.,
RA Wintroub B.O., Rutter W.D., Goldstein S.M.;
RT "Human skin mast cell carboxypeptidase: functional characterization,
RT cDNA cloning, and genealogy."
RL J. Invest. Dermatol. 99:138-145(1992).
RN [4]
RP SEQUENCE OF 110-137.
RX MEDLINE=9214692; PubMed=2708524;
RA Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.O.;
RT "Human mast cell carboxypeptidase. Purification and
RT characterization."
RL J. Clin. Invest. 83:1630-1636(1989).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M27717; AAA35652.1; -.
DR EMBL: M73720; AAA59568.1; -.
DR EMBL: M73716; AAA59568.1; JOINED.
DR EMBL: M73717; AAA59568.1; JOINED.
DR EMBL: M73718; AAA59568.1; JOINED.
DR EMBL: M73719; AAA59568.1; JOINED.
DR EMBL: S40234; AAB22578.2; ALT-SEQ.
DR PIR: A43929; A43929.
DR HSSP: P09955; INSA.
DR MEROPS: M14.010; -.
DR TIM: 114851; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_Carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109
FT CHAIN 110 417
FT METAL 176 176
FT METAL 179 179
FT METAL 304 304
FT ACT_SITE 378 378
FT DISULFID 173 186
FT DISULFID 245 268
SO SPROUNCE 417 AA; 48700 MW; 8CB90DB758117B24 CRC64;

```



```
CC -----
DR EMBL: D78348; BA11366.1; .
DR HSSP: P09955; 1PBA.
DR MEROPS: M14.003; .
DR InterPro: IPR003146; Proper_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Proper_M14; 1.
DR Pfam: PF00246; Zn_carboxypept; 1.
DR PRINTS: PR00765; CARBOXYPTASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109
FT CHAIN 110 416
FT METAL 175 175
FT METAL 178 178
FT METAL 303 303
FT ACT_SITE 377 377
FT DISULFID 244 267
FT DISULFID 258 272
SQ SEQUENCE 416 AA; 47595 MW; DFE1992CD52F8AB4 CRC64;
```

```
alignment_scores:
    Quality: 640.50      Length: 420
    Ratio: 2.390         Gaps: 9
Percent Similarity: 63.810   Percent Identity: 33.810
```

alignment block:

US-09-980-881-1 x CBPB_CANPA ..

Align seg 1/1 to: CBPB_CANPA from: 1 to: 416

```
39 GTCCTGAGGAGCATGTTCTC.....TTCGTGAGGAGCATGTT 76
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
5 I I e l e u v a l t h r l e u a l a l e u a l a s e r l a n i s t y s e r g l y c i n h s . 20
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
77 CTTTCGCGTTCGAGAGTGGCCAGTGTCTAGCTGCTCTTCTAGAACTCTA 126
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
21 .....P h e g l i n g l y c i n u s v a l p h e a r g v a l a s n a l g i u a s p r l u a 35
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
127 GGCAGGTCAAGTTCTTACAGAACTTCTTACATCATATGAGATGTTCTC 176
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
35 s n h i s t i l e a s n l e u l i s t h r l e u a l a s e r t h r t h g i n l i l e a s p h e 51
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
177 TGCAGCGCCGCTAACAGCTGACCTGATTTGTAAGAAAAACAAGTCATTT 226
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
52 T r l u s e r p r o s e r v a l t h r g i n l i l e u s p r o h i s s e r t h r l a l a s p r h 68
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
227 TTTTGTAAATGATCTGATCTGATCTGACAAATGTGAAGCCCATTTAAATGTA 276
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
68 e a r g v a l l y a l a g l a s p r l e u t h r v a l g i u a s p r h e l u s c l n a 85
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
277 GCGGATTCATGACAGTGTCTGCTGCGACAGCTGGAAGATCTTATTCAA 326
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
85 s n o l u e u n i s t y g l u v a l l e u l i l e a s n l e u a r g l e u v a l l e u g l u 101
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
327 CACGAGATTTCACAGACAGCTGACGCCCCGAGCCCTCCGATGCTACTA 376
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
102 G l y l n r h e d l y a r g l n v a l .....P r o l a t h r c l y n h s s e r t y 115
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
377 TGAACAGTATCACTCACTAAATGTAATCTATCTTGATAGAAATTTTAA 426
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
115 r g l u s y t y a s n a r g t r p l u t h r l e g i u a l a t r p t h g l n g l n v a l t 132
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
427 CTGAGAGGATCCGATGATGCTTACAAATCCATGATGATGATCTCATTT 476
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
132 h t e r g l u a n p r o a s p l e u l e s e r t a r g s e r t l e g l y t h r t h p h e 148
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
477 GAGAAGTACCACACTATGATTTTAAAGTTCTTGAGAAACAAACAAACAGC 526
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
149 G l u g l y a r g t h r l e t y r l e u l e u l y s v a l . . . G l y s a l a g l y g l n a s 164
```

```
527 CAAAAATGCCATATGATTCATGCTGGAATTCATGCCAGAAATGATCT 576
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
164 n l y s p r o l a l i l e p h e m e t a s p c y s g l y p h e n i s a l a a r g l u t r p l e s 181
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
577 CTCCTGCTTCTGCTTCTGCTCATTA..... 602
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
181 e r p r o l a p h e t r p g l n t r p h e v a l a r g l u **I l e a r g t h r t y r g l y 197
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
602 ..... 602
198 G l n g l u t l i e n i s m e t h r g l u l e u l e u a s p l y l e u a s p h e t y r v a l l e 214
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
603 .....G E C C A T .....A A T G A A 615
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
214 u p r o v a l g l y a s n l e a s p l y t y r v a l t y t h r t r p t h r l y s a s n a r g m 231
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
616 T G T G A A A A A G A A C C G T T C T T G A T G C G A A C A A T C A T G C A T C G G A A C A 665
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
231 e t t r a r g l y s t h r a r g s e r t h r g l n v a l g l y t h r a s n c y s v a l g l y t h r 247
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
666 G A C T G A A T G C A A C T T T G T C C A A A C A C T G G T G A G A A G G T G A T C 715
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
248 A s p r o t h r a r g a s n p h e . . . A s p a l a g l y t r p c y s l y s l l e g l y a l a s e 263
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
716 C A G T T C C A T G C T C G G A A A C T A C T G T G A C T T T A T C T G A G T C A G A A C 765
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
263 r a r g a s n p r o c y s a s p r l u t h r t y r c y s g l y p r o l a l a l a g l u s e r g l u l 280
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
766 C A G A A G T G A A G G C A G T G G T A G T T C T T G A G A A A A T A T C A A C A G A T T 815
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
280 y s g l u t h r l y s a l a l e u a l a s n p h e l l e a r y s e r a n l e u s e r s e r i l e 296
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
816 A A G C A T A C A T C A G A C A T T C A T T A C C A G A C A T A G T T G A T T C A T A 865
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
297 L y s a l a t y r l e u t h r l i l e n i s s e r t y r s e r g l m e t h e l e u t y r p r o t y 313
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
313 r s e r t y a s p r t y l y s l e u t h r g l u a s n a s n a l a g l u l e u a s n a l a l e u a 330
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
916 C C A G T A A G A G A T G C G T A T T G A C A A A A C T A C T A A A A T A C A G G A T 965
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
330 l a l y s a l a t h r v a l l y s g l u d e u . . . A l a t h r l e u n i s g l y t h r l y s t y r 345
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
966 A C A C A T G C C A T G C C T G A A A C C T A T A C C T A C C T G C G A G G T G G G A 1015
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
346 T h r t y r c l y p r o g l a t a t h r t h r l i l e t y r p r o l a l a l a g l y g l s e r a s 362
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1016 C G A T T G A T C T A T G A T T T G G C A T C A A A T A T T C G T T T A C A T C ..... 1057
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
362 p a s p t r p a l a t y a s p r c l n g l y l e l y s t y s e r p h e t h r p h e l u e u a 379
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1058 .....A A A C C 1063
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
379 r g a s p l y s g l a r g t y r g l y p h e a l a l e u p r o g l u s e r g l n l l e s e r p r o 395
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1064 A C T G T A G A G A A G C T T T G C C G T G C T A A A A T A C T G C A T G C A T G C A T 1113
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
396 T h r c y s g l u g l u t h r l e u l e u a l a l i l e l y s n i s l e u a l a r g t y r v a l l e 412
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1114 T A G A A A T G T T 1123
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
412 u g l n h i s l e u 415
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
seq_name: SwissProt_40:CBPB_PIG
seq_documentation_block:
ID CBPB_PIG STANDARD; PRT; 401 AA.
AC P09955;
DT 01-MAR-1989 (rel. 10, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
```


[illegible]

RA Rees D.C., Lewis M., Lipscomb W.N.;
 RT Refined crystal structure of carboxypeptidase A at 1.54-A
 resolution.";
 RL J. Mol. Biol. 168:367-387(1983).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
 RX MEDLINE-96003618; PubMed-7556081;
 RA Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
 RT "The three-dimensional structure of the native ternary complex of
 RT bovine pancreatic procarboxypeptidase A with propeptidase E and
 RT chymotrypsinogen C.";
 RL EMBO J. 14:4387-4394(1995).
 RN [8]
 RP VARIANT ALLELIC.
 RX MEDLINE-69283620; PubMed-5817619;
 RA Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
 RT "Identification of the amino acid replacements characterizing the
 RT allotypic forms of bovine carboxypeptidase A.";
 RL Biochemistry 8:2762-2768(1969).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
 CC amino acid.
 CC -1- SUBUNIT: MONOMER. THE ZYMOGEN IS SECRETED AS A TERNARY COMPLEX
 CC COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
 CC PROPEPTIDASE E.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -1- DATABASE: NAME-Worthington biochem.com/manual/C/COA.html".
 CC WWW="http://www.worthington-biochem.com/manual/C/COA.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 CC
 DR EMBL; M61851; AAA30426.1; -;
 DR EMBL; M61851; AAA30427.1; -;
 DR EMBL; Z33906; CAB83955.1; -;
 DR PIR; JN0126; CPBA.
 DR PIR; A31406; A31406.
 DR PDB; 3CPA; 15-JAN-87.
 DR PDB; 4CPA; 22-OCT-84.
 DR PDB; 5CPA; 15-JAN-87.
 DR PDB; 6CPA; 15-OCT-91.
 DR PDB; 7CPA; 31-JAN-94.
 DR PDB; 8CPA; 31-JAN-94.
 DR PDB; 1CBX; 31-JAN-94.
 DR PDB; 1CPS; 15-OCT-94.
 DR PDB; 2CTB; 31-JAN-94.
 DR PDB; 1ARL; 01-AUG-96.
 DR PDB; 1ARM; 17-AUG-96.
 DR PDB; 1BAV; 01-APR-97.
 DR PDB; 1YME; 12-FEB-97.
 DR PDB; 1CPX; 05-AUG-98.
 DR PDB; 1PYT; 27-JAN-97.
 DR MEROPS; M14.001; -;
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_carboxept.
 DR Pfam; PF00244; Propep_M14; 1.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
 KW 3D-structure; Polymorphism.
 FT SIGNAL 1 16
 FT PROPEP 17 110 ACTIVATION PEPTIDE.
 FT CHAIN 111 419 CARBOXYPEPTIDASE A.
 FT METAL 179 179 ZINC.
 FT METAL 182 182 ZINC.

FT METAL 306 306 ZINC.
 FT ACT_SITE 358 358 PROTON DONOR.
 FT ACT_SITE 380 380 NUCLEOPHILE.
 FT DISULFID 248 271
 FT VARIANT 289 289
 FT VARIANT 338 338
 FT VARIANT 415 415
 FT CONFLICT 95 95
 FT CONFLICT 199 199
 FT CONFLICT 203 203
 FT CONFLICT 224 224
 FT CONFLICT 232 232
 FT CONFLICT 295 295
 FT TURN 114 114
 FT TURN 119 120
 FT HELIX 125 138
 FT TURN 140 142
 FT STRAND 143 150
 FT TURN 152 153
 FT STRAND 156 162
 FT TURN 171 176
 FT STRAND 180 181
 FT HELIX 183 199
 FT TURN 200 202
 FT HELIX 204 212
 FT STRAND 214 218
 FT HELIX 223 231
 FT TURN 232 232
 FT TURN 234 235
 FT STRAND 242 242
 FT TURN 244 245
 FT STRAND 249 249
 FT TURN 253 254
 FT TURN 261 262
 FT STRAND 267 267
 FT TURN 270 271
 FT TURN 273 274
 FT STRAND 275 275
 FT TURN 280 281
 FT HELIX 284 296
 FT STRAND 299 307
 FT TURN 311 314
 FT STRAND 314 314
 FT TURN 324 325
 FT HELIX 326 344
 FT STRAND 349 352
 FT HELIX 353 356
 FT TURN 357 357
 FT HELIX 364 370
 FT TURN 371 372
 FT STRAND 375 381
 FT TURN 388 389
 FT HELIX 393 416
 SQ SEQUENCE 419 AA; 47082 MW; 21B86407B3BRC452 CRC64;

alignment_scores:
 Quality: 536.00 Length: 431
 Ratio: 2.144 Gaps: 12
 Percent Similarity: 58.005 Percent Identity: 32.715

alignment_block:
 US-09-980-881-1 x CPBA_BOVIN ..

Align seg 1/1 to: CPBA_BOVIN from: 1 to: 419

24 CTTTCAGCCTTCAGATCTTACCCATTGTTCTTCTGACGACGA 73
 ||| |||:|||||:|:|:| |||:|
 4 LeuLeuLeuSerValLeuLeuGlyAlaAlaLeuGlyLysGluAsp.. 19
 74 TGTCTTCGGCTTCAGAGTGGCCAGTTCTGCTGCTTCTAGACCT 123
 ||| :|:| ||||| |||
 20PheValGlyHisGlnValLeuArgIleThrAlaAlaAspG 33

```

124 CTAGCAAGTTTCAGTTCTACAGAACTTATACAAACATATGAGATTGT 173
   :: ::::::::::::::::::::
33 LIALGLVALGLINThrValLysGLuLeuGLuAspLeuGLuHISLeuGLu 49
174 CTC.....TGG.....CAGCGGTAACAGCTGACCTTATTGT 205
   ||| ||| |||
50 LeuAspPheTrpArgGLyProGLyGlnPro..... 59
206 GAAGAAACAAACAGTCCATTTTGTAAATGATCATCTGATGTCGACAA 255
   :: :: ::|
60 ... GlySerProIleAspValArgValProPheProSerLeuGlnAla 75
256 TGAAGACCCATTTAAATGTCAGCGGAATTCATGACAGTCTTGTGCGCA 305
   ||| ::| ::|
75 AltyValAlaPheLeuGlnAlaHISGLyIleArgTyrArgIleMetIleGlu 91
306 GACGTGGGAAGATCTTATT.....CAACAGAGATTTCACACACAC 346
   ||| ::| ::|
92 AspValGlnSerLeuLeuAspGluGlnGlnGlnGlnMetSerLeuAla 108
347 AGTCAGCCCCGAGCTCCGATCGTAC...TATGAACAGATTACCTACAC 393
   ||| ::| ::|
108 nSerArgAlaArgSerThrAsnThrPheAsnTyrAlaThrTyrHISThr 125
394 TAAATGAAAATCTATTCTTGATAGAAATTTATACTGAGAGGCACTGAT 443
   ||| ::| ::|
125 euAspIleuIleTyrAspPheMetAspLeuValAlaGlnHISProGln 141
444 ATGCTTACAAAATCCATGATGATGCTCATTTGAGAAATGCCACTGTA 493
   ::| ::| ::|
142 LeuValSerLysLeuGlnIleGlyArgSerTyrGluGlyArgProIle 158
494 TGTTTTAAAGTTTCTGAAAAAGAACACAGCCAAAATGCCATATGCA 543
   ||| ::| ::|
158 rValLeuLysPheSer...ThrGlySerAsnArgProAlaIleTrp 174
544 TTGACTGTGGATCCATGCCACAGAAATGATCTCTCTGTTCTGCTTG 593
   ||| ::| ::|
174 LeuAspLeuGlyIleHISSerArgGluTrpIleThrGlnAlaThrGly 190
594 TGGTTCATA..... 602
191 ThrPheAlaLysLysPheThrGluAspTyrGlyGlnAspProSerPhe 207
602 ..... 602
207 rAlaIleLeuAspSerMetAspIlePheLeuGlnIleValThrAsnPro 224
603 .....GCCCATATCGAATGTGGAGAAAGAACCGT 632
224 spGlyPheAlaPheThrHISSerGlnAsnArgLeuTrpArgLysTrpArg 240
633 TCTTCTATGCGAACAATCATCTGATCGAAGACAGCTGAATGCAACT 682
   ||| ::| ::|
241 SerValThrSerSerSerLeuGlyValAlaAspAlaAsnArgAsnTr 257
683 TGTCTCCAAACACTGGTGTGAGAGAGTGATCCATCTCATGCTCGG 732
   :: :: ::|
257 P...AspAlaGlyPheGlyLysAlaGlyAlaSerSerProCysSerG 773
773 AAACCTACCTGAGCTTTATCTGACTCAGAACACAGAGAGGAGCAGTG 782
   ||| ::| ::|
273 LThrTyrHISGLyLysTyrAlaAsnSerGluValGlnValLysSerIle 289
783 GCTAGTTCTTGAGAAATATCATCAGATTAAGCATATCATGAGCAT 832
   ::| ::| ::|
290 ValAspPheValLysAspHISGLyAsn...PheLysAlaPheLeuSer 305
833 GCATTCTATCTCCAGCATATAGTGTTCATATTCATATACAGAAATA 882
   ||| ::| ::|
305 eHISSerTyrSerGlnLeuLeuLeuTyrProTyrGlyTyrThrThrGln 322
883 AAAGCAAAAGACCATGAGAACTGTCTACTAGCCACTGAAGAGTTGCT 932

```

```

:: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
322 eHISLeuProAspLysThrGlnLeuAsnGlnValAlaLysSerAlaValGlu 338
933 GCTATTGACAAAACATAGTAAATATACAGATATACATGCGCATGCGTC 982
   ||| ::| ::| ::|
339 AlaLeu...LysSerLeuTyrGlyThrSerTyrLysTyrGlySerIle 354
983 AGAAACCTTATACCTGATCTCTGAGAGTGGGAGCATGATGATCATGAT 1032
   ||| ::| ::|
354 eHISThrIleTyrGlnAlaSerGlyLysSerIleAspTrpSerTyrAsnG 371
1033 TGGGCAATCAAAATTCGTTTACATCAA..... 1060
   ||| ::| ::|
371 IndIleLysTyrSerPheThrPheGluLeuArgAspThrGlyArgTyr 387
1061 .....CCGACCTGAGAGAGCTTT 1080
   ||| ::| ::|
388 GlyPheLeuLeuProAlaSerGlnIleIleProThrAlaGlnIleThr 404
1081 TGGCGCTGTCTTAAATAGCTTGCAATGATCATTAAGAAATGTT 1123
   ::| ::| ::|
404 pLeuGlyValIleuThrIleMetGlnHISThrLeuAsnAsnLeu 418
seq_name: SwissProt_40:CBPL_RAT
seq_documentation_block:
ID CBPL_RAT STANDARD: PRT; 419 AA.
AC P00731;
DT 21-JUL-1986 (Rel. 01, Created)
DI 01-FEB-1996 (Rel. 33, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN CPAL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82105986; PubMed=6275388;
RX Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standing D.N.,
RA Picket R.L., Valenzuela P., Rutter W.J.;
RT "Rat procarboxypeptidase A: cDNA sequence and preliminary
RT characterization of the gene."
RT Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family."
RL J. Biol. Chem. 263:17837-17845(1988).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01232; CAA24542.1; -
DR EMBL; J00713; AAA40893.1; -
DR EMBL; M23990; AAA40955.1; -
DR EMBL; M23960; AAA40955.1; JOINED.
DR EMBL; M23985; AAA40955.1; JOINED.
DR EMBL; M23986; AAA40955.1; JOINED.

```



```

684 GTTCCAAACACCTGCTGAGAGAGATCCAGTTCCTCATCTCGCA 733
    ::::::::::::::::::::::::::::
256 ..AspIaagIyPheGlyGlyProGlyAlaSerSerInProCysSerAs 271
734 AACCTAGCTGACTTATCTGATGACAGAACCAAGTGAAGCAGCAG 783
    ::::::::::::::::::::
271 PSETTYHISGLYProSerAlaAsnSerGluValGluValIlySerIleV 288
784 STAGTCTTGTGAGAGAAATATCAACAGATTAAGCATATCATAGCAT 833
    ::::::::::::::::::::
288 AlaPheIleIleYSerHis...GlyValIlySerAlaPheIleIleu 303
834 CATTCAGTCTCCAGCATATAGTCTTCCATATTCCTATACAGCAAGTAA 883
    ::::::::::::::::::::
304 HisSerYSerGlnIleuMetPheProGlyGlyIlySerCysThrIly 320
884 AAGCAAGACCATGAGCAAGCTCTAGTACGAGCAAGCAAGCATCTCTG 933
    ::::::::::::::::::::
320 SLeuAspAspPheAspGlnIleuSerGluValAlaGlnIlyAlaGln 337
934 STATTGCAAAAGTAAATAATACAGATATACATGAGCATGAGCTCA 983
    ::::::::::::::::::::
337 erIleu...ArgSerLeuHisGlyThrIlySerValGlyProIleCys 352
984 GAAACSTTATACSTAGCTCTGAGAGTGGGAGCATGGATCTATGATTT 1033
    ::::::::::::::::::::
353 SerValIleYSerGlnAlaSerGlySerIleAspTrpSerYlyAspTy 369
1034 GCGCATCAAAATATCTGTTACATCAAAC..... 1061
    ::::::::::::::::::::
369 rGlyIleYSerPheAlaPheGlnIleuArgAspThrGlyArgTyG 386
1062 .....CCACSTGTAGAGAGCTTTGCGC 1085
    ::::::::::::::::::::
386 IyPheIleuLeuProAlaArgGlnIleuPro 396
seq name: SwissProt_40:CBPL_HUMAN

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

DR EMBL; X67318; CAA47732.1; -.
DR PIR; S02810; S02810.
DR PIR; S29127; S29127.
DR HSSP; P00730; 1PPT.
DR MEROPS; M14.001; -.
DR MTM; I14850; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_Carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_Carboxypept; 1.
DR PRINTS; PR00765; CARBOXYPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR Hydrolase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT STGNPL 1 16
FT PROPEP 17 110 ACTIVATION PEPTIDE.
FT CHAIN 111 419 CARBOXYPEPTIDASE A1.
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 182 182 ZINC (BY SIMILARITY).
FT METAL 306 306 ZINC (BY SIMILARITY).
FT ACT_SITE 358 358 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 380 380 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 248 271 BY SIMILARITY.
SQ SEQUENCE 419 AA; 47140 MW; 439FAFFAE858B1 CRC64;

alignment_scores:
    Quality: 489.50      Length: 401
    Ratio: 2.128        Gaps: 13
    Percent Similarity: 57.357      Percent Identity: 32.668

alignment_block:
US-09-980-881-1 x CBPL_HUMAN ..

Align seg 1/1 to: CBPL_HUMAN from: 1 to: 419

24 CTTTCAGACCTTGACGCTTGTACCCATGTTCTCTCTGAGCAGCA 73
    ||| ::::::::::::::::::::
4 LeuLeuValIleSerValIleuLeuGlyAlaValPheGlyIlyGluAsp.. 19
74 TGTCTTCCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCT 123
    || ::::::::::::::::::::
20 .....PheValGlyHisGlnValIleuArgIleSerValAlaAspG 33
124 CTAGCAGATTCAGTCTCTCAGAAATCTTACTACAAATATGAGATTGTT 173
    ::::::::::::::::::::
33 IuAlaGlnValGlnIlySerValIlySerGluAspLeuGlnHisIleuGln 49
174 CTC.....TGGCAG.....CCGGTAACAGCTGA 196
    ||| ::::::::::::::::::::
50 LeuAspPheIlePheGlyProAlaHisProGlySerProIleAspVal.. 65
197 CTTATTGTGAAGAAAAACAAGTCCATTTTTGTAAATGCAATGCTGATG 246
    ::::::::::::::::::::
66 .....ArgValProPhe.....ProSerI 72
247 TCGCAATGTGAAGCCCATTTAATGTGAGCGGCAATTCATGAGTGTC 296
    ::::::::::::::::::::
72 IeGlnAlaValIlyIlePheLeuGlnSerHisGlyIleSerTyGluThr 88
297 TTGTCGCGACAGCTGGAAGATCTTATT.....CAACAGCAGATT.. 335
    ::::::::::::::::::::
89 MetIleGluAspValGlnSerIleuLeuAspGlnIlyGlnIleuPhe 105
36 .....TCCACGACACAGTCAAGCCCCGAGGCT 363

```



```

54 rpylserproserSerPheasnArgProValAspValLeuValProSer 70
    ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
213 AAACAAGTCCATTTTGTAAATGCATCTGANTGCACAATGGAAGC 262
    ::::: ||| ::::: ||| ::::: ||| ::::: |||
71 ValSerLeuValLaphe.....LysSe 78
263 CCATTAAATGTAGCGGAATTCACAGTCTCTGTCGACAGCTGG 312
    : ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
78 rPheLeuArgSerGlnGlyLeuGluValValThrIleGluAspLeug 95
313 AAGATCTTAT.....CAACAGAGATTCCACAGACACAGCTGAC 353
    :: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
95 lnaLaleuLeuAspAsnGluAspAspGluMetGlnHisAsnGluGln 111
354 CCCCAGGCTCCGACATCGTAC...TATGACAGATGATCAGTAAATGA 400
    ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
112 GluArgSerSerAsnAsnPheAsnGlyAlaValThrHisSerLeuGluAl 128
401 AATCTATTCTTGATAGAAATTATTAAGTACGAGGACATCTGATATGCTTA 450
    ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
128 aileTyHisGluMetAspAsnIleAlaAlaAspPheProAspLeuAla 145
451 CAATAATCCATGTGATCTCTCATTTGAGAAAGTACCCACCTATGTTTA 500
    ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
145 rGaTyValLysIleGlyHisSerPheGluAsnArgProMetTyValLeu 161
501 AAGCTTCTGGAAGAAACAACAAGCCAAATAATGCATATGATGATGACTG 550
    ||| ||| ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
162 LysPheSerThrGlyLysGlyValArgArgProAlaValThrLeuAsnAl 178
551 TGGAAATCCATGCCAGAGATGATCTCTGCTTCTGCTTCTGCTGG... 596
    ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
178 acGlyIleHisSerArgIleuTrpIleSerGlnAlaThrAlaIleTrpHra 195
596 ..... 596
195 laArgLysIleValSerAspTyGlnArgAspProAlaIleThrSerIle 211
597 .....TTCATA..... 602
212 LeuGluLysMetAspIlePheLeuProValAlaAsnProAspGlyTy 228
603 .....GCCATAATGGAATGTGGAGAAAGAACCGTTCTTCTTCT 639
228 rValTyThrGlnThrGlnAsnArgLeuTrpArgLysThrArgSerArga 245
640 ATGCGAACAATCATGTGCATGCGAAGACCTGATACCACTTGTGCTCC 689
245 snProGlySerSerCysIleGlyAlaAspProAsnArgAsnTrp...Asn 260
690 AAACACTGCTGTGAGGAAGCTGCATCTCTGATCTCGGAAACCTA 739
261 AlAserPheAlaGlyLysGlyAlaSerAspAsnProCysSerGlyValTy 277
740 CTGTGACTTATCTCGAGTCAGACAGAACCGAGAGGAGCTGCTGTT 789
277 rHisGlyProHisAlaAsnSerGlyValGluValLysSerValValAsp 294
790 TCTTGAGAAAGAAATTCACAGATTAAACATATACATGACATGACATCA 839
294 heIleGlnLysHisGlyAsn...PheLysGlyPheIleAspLeuHisSer 309
840 TACTCCAGCATATAGTGTTCATATCTCATATACAGAACTAAAGCAA 889
310 TySerGlnLeuLeuMetTyProTyGlyTySerValLysLysAlaArg 326
890 AGACCATGAGGAAGTCTCTAGTACGACAGTGAAGCGTTCTGCTATTG 939
326 oAspAlaGluGlnLeuAspLysValAlaArgLeuAlaAlaLysAlaLeu 343
940 ACAAACTAGTAAATAATACAGATATACATGCGCATGGCTCGAAGACC 989
    ::::: ||| ::::: |||

```

```

343 laSerValSer...GlyThrGluTyGlnValGlyProThrCysThrThr 358
990 TTATACCTAGCTCTCGAGTGGGACAGATGGATGATGATTTGGGCAAT 1039
    ::::: ||| ::::: ||| ::::: ||| ::::: |||
359 ValTyProAlaSerGlySerSerIleAspTrpAlaTyAspAsnGlyTl 375
1040 CAATATTCGTTTACA 1055
    ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
375 elysPheAlaPheThr 380

```

seq_name: SwissProt_40:CBPZ_SIMV1

seq_documentation_block:

```

ID CBPZ_SIMV1 STANDARD; PRT; 304 AA.
AC P42788;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc carboxypeptidase (EC 3.4.17.*) (Fragment).
OS Simulium vittatum (Black fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Simuliidae; Simulium.
OX NCBI_TaxID=7192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cut;
RX MEDLINE=94093864; PubMed=8269093;
RA Ramos A., Mahowald A., Jacobs-Lorena M.;
RT "Gut-specific genes from the black fly Simulium vittatum encoding
RT trypsin-like and carboxypeptidase-like proteins."
RL Insect Mol. Biol. 1:149-163(1993).
CC -1- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
CC -1- TISSUE SPECIFICITY: GUT-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L08481; AAL18531.1; -.
CC DR HSSP: P48052; IAYE.
CC DR InterPro: IPR000834; Zn_carboxpept.
CC DR Pfam: PF00246; Zn_carboxpept. 1.
CC DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
CC DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
CC KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc.
CC FT NON_TER 1
CC FT METAL 58
CC FT METAL 61
CC FT METAL 184
CC FT ACT_SITE 236
CC FT ACT_SITE 259
CC FT DISULFID 125
CC FT DISULFID 148
CC SQ SEQUENCE 304 AA; 34849 MW; 2E6E3FF8A6A9144 CRC64;

```

alignment_scores:

```

Quality: 409.00 Length: 308
Ratio: 2.260 Gaps: 10
Percent Similarity: 58.766 Percent Identity: 31.818

```

alignment_block:

US-09-980-881-1 x CBPZ_SIMV1 ..

Align seg 1/1 to: CBPZ_SIMV1 from: 1 to: 304

381 CAGTATACCTGCTAAATGAATCTATTTCTTGATAGATTATTAAGTGA 430

OM of: US-09-980-881-1 to: SPTREMBL_19:* out_format : pfs
Date: Sep 18, 2002 4:42 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+g2p_model -DEV=xlh
-O=/cgn2.1/USPTO_sptool/US09980881/runatc_16092002_140041_8910/app-query.fasta_1.1655
-DB=SPTREMBL_19 -OFT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPTOL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAEXT=7.000 -TGAPOP=10.000 -TGAEXT=0.500
-DELDP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsnum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFT=pts
-NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881.ecgnl_1.101 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-1
Query length: 1573
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 17294929
Search time (sec): 93.320000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Document
sp_human:Q15114	1885.00	3342.23	1.9e-178	423	Q15114 homo sapiens (human). pc
sp_human:Q9PZY6	1884.00	3342.01	2.3e-178	360	Q9PZY6 homo sapiens (human). ca
sp_human:Q961Y4	1880.00	3333.31	6.0e-178	423	Q961Y4 homo sapiens (human). ca
sp_rhodent:Q9JHM6	1592.50	2820.48	2.2e-149	422	Q9JHM6 mus musculus (mouse). ca
sp_rhodent:Q9QZF0	1586.50	2809.77	5.4e-147	422	Q9QZF0 mus musculus (mouse). ca
sp_rhodent:Q9E0V9	1568.50	2777.66	8.4e-147	422	Q9E0V9 rattus norvegicus (rat). ca
sp_human:Q9NFI8	1029.00	1822.61	1.8e-93	198	Q9NFI8 homo sapiens (human). da
sp_vertebrate:Q9PUF2	714.00	1253.50	4.3e-62	416	Q9PUF2 bothrops jararaca (jarar
sp_human:Q96B94	652.50	1143.77	5.6e-56	417	Q96B94 homo sapiens (human). si
sp_human:Q96B08	646.00	1132.20	2.5e-55	416	Q96B08 homo sapiens (human). hy
sp_mammal:Q9ASR3	642.50	1126.05	5.5e-55	412	Q9ASR3 sus scrofa (pig). proca
sp_rhodent:Q975Y7	531.00	930.93	5.9e-44	279	Q975Y7 rattus norvegicus (rat). 22
sp_rhodent:Q9CVD1	516.50	901.12	1.8e-42	419	Q9CVD1 mus musculus (mouse). 22
sp_mammal:Q9TVH85	511.00	891.35	6.4e-42	417	Q9TVH85 sus scrofa (pig). carbox
sp_human:Q96ON3	511.00	891.35	6.4e-42	417	Q96ON3 homo sapiens (human). ca
sp_human:Q96A12	489.50	852.95	8.7e-40	419	Q96A12 homo sapiens (human). hy
sp_human:Q9B867	489.50	852.95	8.7e-40	419	Q9B867 homo sapiens (human). ca
sp_human:Q9B829	484.00	848.27	2.7e-39	247	Q9B829 homo sapiens (human). ca
sp_invertebrate:Q9VL87	460.00	800.21	7.5e-37	424	Q9VL87 drosophila melanogaste
sp_invertebrate:Q9W475	459.00	788.40	1.2e-36	1192	Q9W475 drosophila melanogaste
sp_invertebrate:Q9VL42	451.50	785.10	5.2e-36	422	Q9VL42 drosophila melanogaste
sp_invertebrate:Q9VH21	436.00	754.30	2.0e-34	584	Q9VH21 caenorhabditis elegans
sp_invertebrate:Q9W478	418.00	724.93	1.1e-32	440	Q9W478 drosophila melanogaste
sp_invertebrate:Q9V421	407.50	706.68	1.2e-31	419	Q9V421 drosophila melanogaste
sp_invertebrate:Q9V566	402.00	696.75	4.3e-31	424	Q9V566 drosophila melanogaste
sp_invertebrate:Q9VH86	402.00	696.61	4.3e-31	430	Q9VH86 drosophila melanogaste
sp_invertebrate:Q9VH22	400.50	691.73	6.5e-31	540	Q9VH22 caenorhabditis elegans
sp_invertebrate:Q9VH32	399.50	692.22	7.7e-31	427	Q9VH32 aedes aegypti (yellowf
sp_invertebrate:Q19121	397.00	690.47	1.3e-30	323	Q19121 caenorhabditis elegans
sp_invertebrate:Q9V422	396.00	685.41	1.7e-30	453	Q9V422 drosophila melanogaste
sp_invertebrate:Q9V423	386.50	669.12	1.5e-29	423	Q9V423 drosophila melanogaste
sp_invertebrate:Q9V424	382.50	658.91	4.0e-29	581	Q9V424 caenorhabditis elegans
sp_invertebrate:Q9V425	370.00	628.77	7.9e-28	1430	Q9V425 drosophila melanogaste
sp_invertebrate:Q9N3S6	370.00	637.23	7.0e-28	545	Q9N3S6 caenorhabditis elegans
sp_invertebrate:Q9VCM8	368.50	636.05	9.4e-28	467	Q9VCM8 drosophila melanogaste
sp_invertebrate:Q9V426	362.50	623.60	3.7e-27	455	Q9V426 caenorhabditis elegans
sp_invertebrate:Q9XU75	358.00	613.88	1.1e-26	666	Q9XU75 caenorhabditis elegans
sp_invertebrate:Q9V564	355.00	613.12	2.0e-26	415	Q9V564 drosophila melanogaste
sp_invertebrate:Q9V567	354.50	612.58	2.2e-26	400	Q9V567 drosophila melanogaste
sp_invertebrate:Q9V568	344.50	595.93	2.1e-25	354	Q9V568 drosophila melanogaste

sp_invertebrate:Q9V428 + 344.50 594.32 2.2e-25 418 | Q9V428 drosophila melanogaste
sp_invertebrate:Q9V429 + 337.00 583.77 1.2e-24 312 | Q9V429 drosophila melanogaste
sp_human:Q9NFI9 + 334.00 578.33 2.3e-24 315 | Q9NFI9 homo sapiens (human). ca
sp_vertebrate:Q9QW12 + 324.50 567.89 1.7e-23 161 | Q9QW12 brachydanio rerio (ze
sp_fungi:Q14418 + 320.00 550.61 6.1e-23 418 | Q14418 metarhizium anisoplia

seq_name: sp_human:Q15114

seq_documentation_block:

ID: Q15114 PRELIMINARY; PRT; 423 AA.
AC: Q15114;
DT: 01-NOV-1996 (TRENBERG, 01, Created)
DT: 01-NOV-1996 (TRENBERG, 01, Last sequence update)
DT: 01-DEC-2001 (TRENBERG, 19, Last annotation update)
DE: PCPB PROTEIN.
GN: PCPB.
OS: Homo sapiens (Human).
OC: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC: Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX: NCBI_TaxID=9606;
RN: [1]
RP: SEQUENCE FROM N.A.
RC: TISSUE=LIVER.
RX: MEDLINE-92042093; PubMed-1939207;
RA: Eaton D.U., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT: "Isolation, molecular cloning, and partial characterization of a novel
RT: carboxypeptidase B from human plasma."
RL: J. Biol. Chem. 266:21833-21838(1991).
DR: EMBL: M75106; AAA60042.1; -
DR: HSSP: P00730; 1PRT.
DR: MEROPS; M14.009; -
DR: InterPro; IPR003146; Propep_M14.
DR: Pfam; PF02244; Propep_M14; 1.
DR: Pfam; PF00246; Zn_carboxypept. 1.
DR: PRINTS; PR00765; CRBOXPTASEA.
FT: CHAIN: 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ: SEQUENCE 423 AA; 48442 MW; 9B38327F6E79E4 CRC64;

alignment_scores:

Quality: 1885.00 Length: 423
Ratio: 5.122 Gaps: 2
Percent Similarity: 86.998 Percent Identity: 86.525

alignment_block:

US-09-980-881-1 x Q15114 ..
Align seg 1/1 to: Q15114 from: 1 to: 423

18 ATGAAGCTTTGAGCGCTTGCAGTCCCTTGACCATGTTCTCTCTGCA 67
|||||
1 MetylstleucyssetleuAlaValleuValprolleValleuPhcysgl 17
68 GCAGCATGCTTTCGCGTTCACAGTGGCCCAAGTTCTAGCTGCTTCCGA 117
|||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValleuAlaValleuProA 34
118 GAACCTTAGGCAAGTCAAGTCTTACGAATCACTTACATCAACATATAG 167
|||||
34 rGThHisSerArgGlnValGlnValleuGlnAsnLeuThrThrTyGlu 50
168 ATTGTTCTCTGCGACCGGTAAACAGTCACTTATGTAAGAAAAACA 217
|||||
51 lleValleuIleuTrpGlnProValThrAlaSplenuIleVallyGlyGyl 67
218 AGTCCATTTTGTGAATGCAATGATGTCAGATGCAATGTGAAGCCCAT 267
|||||
67 nValHisPhePheValAsnAlaSerAspValAlaSplenuIleVallyGyl 84
268 TAAATGTAGCGGAATTCAGTCACTGCTGCTGCGACAGTGAAGAT 317
|||||
84 euAsnValSerIlePheProCysSerValleuValleuAlaSplenuSpl 100

```

318 CTTATTCAACAGACAGATTTCCACAGACAGTCCAGCCCCGAGCCTCCGC 367
|||||
101 leuileglnnglnlnlleserbsnspthrsalserproarglaserall 117
368 ATCTACTATGAGACAGTACTACTAAATGAATCTATCTTGAGAG 417
|||||
117 aserlytrgluglnltnrhlsserleusnolletylertrpilleg 134
418 AATTTATAGTAGAGAGCATCCGTATATGCTTACAAAAATCCACATGGA 467
|||||
134 lupheillethrcpluarghlsproaspmetleuthrlysllehlsllegly 150
468 TCCCATTTGAGAGTACCACACTATATGTTTAAAGTTTGGAAAAAGA 517
|||||
151 SerSerPheglulystyrproleuylvalleuylvalserglylysgl 167
518 ACAAACAGCAAAAAATGCCATATGAGTACGATGTCGTAATCCAGCCAG 567
|||||
167 uclnthralalysasnalailetrpilaaspccyllehlslalarg 184
568 AATGAGATCTCTGCTTCTGCTTGTGTCATAGGCCAT ..... 608
|||||
184 lurtplleserproalabhecysleutrpheilleglyhlsllethrgln 200
608 ..... 608
201 phefyrglylleilleglylntyrthralsnleuuargleuvalasprh 217
608 ..... 608
217 eTyrValmetProvalValasnsvalasprglytyrasprlysertrpysl 234
609 ..AATCGAATGTGAGAAAGAACCGTCTTTCTATGCGAACAATCATATGC 656
|||||
234 ysaanaargmettrparglysaanaargserpheyralaasnaahnlscys 250
657 ATCGAAGACAGCTGAATAGCACTTTGTCSCAAACACGTCGTGAGGA 706
|||||
251 lleelllythrspheusnargasnphelalaserlyshlstrpcysglugl 267
707 AGGTGATCCAGTTCCTCATGCTCGGAACCTACTGAGATTATCTCTG 756
|||||
267 ucllylaserseerSerSerlyserglutntrlycysgllyeutytrprog 284
757 AGTAGAACAAGAGTAGAGCAGTGCCTAGTCTTGAGAAAGATATC 806
|||||
284 lusergluprogluvallysalvalalaserpheleuarargasnille 300
807 AACGAGATTAAAGCATATCAGCATGATTCATCTCCAGCATATAGT 856
|||||
301 AsnqlnlelysalatylrlleserMetHlserlyserqlnHlslleval 317
857 GTTTCATATTCCTATACAGCAAGTAAACCAAGACCATGAGAACTGT 906
|||||
317 lPhePTySerytrThrargSerlySerlyAsprHlsgluldeus 334
907 CTTAGAGGACAGTAGAGCAGTCTGCTATGTGACAAACTAGTAAAT 956
|||||
334 erleuvalalaserglualalvalalleglulsthrSerlyAsn 350
957 ACCAGATATACATGCGCATGGCTCAGAAACCTTATACCTAGTCTCTG 1006
|||||
351 ThrArglytrhrnlsglyhlsglyserglutntrleuylrleuualarprogl 367
1007 AGGTGGGAGCATTTGATCTATGATTTGGGCATCAATATTCGTT ..... 1051
|||||
367 yclglylAspAsprlrlelytrAsprleuglylleystyrSerPheThrI 384
1052 .....TAC 1054
384 lecluleuarlgasprhnglythrttyrlypheleuLeuProgluarglytr 400

```

```

1055 ATCAAAACCCACCTGTAGAGAGCTTTTGCCGCTGCTCTATAAATAGCTTG 1104
|||||
401 lleuysProthrCysArgglualalphehlalalvalserlyslleallatr 417
1105 GCARGTCTATTAGGAATGTT 1123
|||||
417 phlsvallelcargsnval 423
seq_name: sp_human:Q9P2Y6

```

```

seq_documentation_block:
ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B-LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL: AB011969; BAA90475.1;
DR HSSP: P00730; 2CTC.
DR MEROPS: M14.009;
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxept; 2.
DR PRINTS: PR00765; CRHOXYPTASEA.
SQ SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

```

```

alignment_scores:
Quality: 1884.00 Length: 360
Ratio: 5.263 Gaps: 0
Percent Similarity: 99.444 Percent Identity: 98.333

```

alignment_block:

```

US-09-980-881-1 x Q9P2Y6 ..
Align seg 1/1 to: Q9P2Y6 from: 1 to: 360

```

```

18 ATAGAGCTTTGACAGCTTGACGTCCTTTGACCATGTTGCTCTGTGTA 67
1 MetLysLeucysSerleuAlaValleuValProIlleValleuPheCysgl 17
68 GCAGCATGTCCTGCGCTCCAGAGTGCGCAAGTTCTAGCGCTCTCTTA 117
|||||
17 uclnlnhsvalalphealalpheglntntrglylnvalleuAlalaleuProA 34
118 GAACCTCTAGGCAAGTTCAAGTCTACAGAACTTACTACACATATGAG 167
|||||
34 rghrseraraglnvalglvalleuglnasnleuthrthrlytrglu 50
168 ATGTCTCTGTGGACCGGTAACAGCTGACCTTATTTGTAAGAAAAACA 217
|||||
51 llevalleutrpelaprovalthralaasprleuIllevallyllysgl 67
218 AGTCATTTTGTAAATGCAATGATGTCGACATGTAAGTGAAGGCAT 267
|||||

```



```

67 nValHisPhePheValAsnAlaSerAspValAlaSerValLysAlaHisL 84
268 TAAATGTGACGGGAATTCATGACAGTGTCTGTGTCGACAGCGTGAAGAT 317
      |||
84 euAsnValSerGlyIleProCysSerValLeuAlaAspValGlnAsp 100
      |||
318 CTATTATCAACAGCAGATTTCACACAGACAGTACGCCCCGAGCCCTCCG 367
      |||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaSerAl 117
      |||
368 ATGCTACTATGACAGATATCATCTACATAAATGAAATCTATCTGTGATAG 417
      |||
117 aSerTyrTrpGlnGlnIleThrHisSerLeuAsnGlnIleTyrSerTrpIleG 134
      |||
418 AATTATATACAGAGCAGCATCTGATATGCTTACAAAATCCACATTTGA 467
      |||
134 lUpheIleThrGlnAlaGHisProAspMetLeuThrLysIleHisIleGly 150
      |||
468 TCCATATTGAGAAAGTACCCACCTATATGTTTAAAGTTTCTGAAAAGA 517
      |||
151 SerSerPheGlnLysTyrProLeuTyrValLeuLysValSerGlyLysG 167
      |||
518 ACAACAGCCAAAATGACCATATGATGATGACTGTGGAATCCATGCCAGAG 567
      |||
167 uGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
      |||
568 AATGATCTCTCTGCTTCTGCTTGTGCTTGTATGATGAGCCATATATGAA 617
      |||
184 lUrTrpIleSerProAlaPheCysLeuThrPheIleGlyHisAsnAlaArgMet 200
      |||
618 TGCAGAAAGAACCGTTTCTTATATGCAACATATGATGATGCAAGACA 667
      |||
201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAs 217
      |||
668 CCTGATATGACAACTTGTCTCCAAACACTGCTGTGAGGAAGAGTGCATCCA 717
      |||
217 PheAsnArgAsnPheAlaSerLysHisTrpCysGlnGlnLysAlaSer 234
      |||
718 GTTCCATGCTCGGAAACCTACTGTGACTTATCTGATGATGACAGACA 767
      |||
234 eSerSerCysSerGlnThrTyrCysGlyLeuTyrProGlnSerGlnPro 250
      |||
768 GAATGAGCAGCAGTGGCTAGTTTCTTGAAGAAATATCAACAGATTA 817
      |||
251 GlnValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIle 267
      |||
818 AGCATATACAGCATGATCATCTACCTCCAGCATATAGTGTTCATAT 867
      |||
267 sAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 284
      |||
868 CTTATACAGCAAGTAAAGCAAGACCATGAGAACTGTCTAGTAGCC 917
      |||
284 eTyrTrpThrArgSerLysSerLysAspHisGlnLeuSerLeuValAla 300
      |||
918 AGTGAACAGATTCGTCTATTGACAAAACCTAGTAAATACAGATATAC 967
      |||
301 SerGlnAlaValAlaArgAlaIleGlnLysIleSerLysAsnThrArgTyr 317
      |||
968 ACATGAGCCATGAGTCAAGAACTTATACCTGATCTGAGAGTGGGAGCG 1017
      |||
317 rHisGlyHisGlySerGlnThrLeuTyrLeuAlaProGlyGlyLysAsp 334
      |||
1018 AATTGATCTATGATTGGGACATCAAAATATCTTACATCAAAACCACT 1067
      |||
334 sPrTrpIleTyrAspLeuGlnIleLysTyrSerPheThrSerAsnProPro 350
      |||
1068 GTAGAGAAAGCTTTTGGCGCTCTCTATAA 1097
      |||
351 ValGlnLysLeuLeuProLeuSerLeuLys 360

```

seq_name: sp_human:Q961Y4
seq_documentation_block:

```

ID 0961Y4 PRELIMINARY; PRT; 423 AA.
AC 0961Y4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B2 (PLASMA).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKETETAL MUSCLE;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC007057; AAH07057.1; -.
SQ SEQUENCE 423 AA: 48412 MW: 9B383E03F6E7CF5 CRC64;

```

alignment_scores:

Quality: 1880.00 Length: 423
Ratio: 5.109 Gaps: 2
Percent Similarity: 86.998 Percent Identity: 86.288

alignment_block:

us-09-980-881-1 x 0961Y4 ..

Align seg 1/1 to: 0961Y4 from: 1 to: 423

```

18 ATGAAGCTTTCAGACCTTTCAGAGTCTTGTACCATTTGTTCTTCTGTGA 67
      |||
1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysG 17
      |||
68 GCAGCATGCTTGGGCTTCAGAGTGGCCAGTGTACTGCTGCTCTCTA 117
      |||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuPro 34
      |||
118 GAACCTCTAGCCAAATTCAAGTTCTACAGAACTTACTACACAATATGAG 167
      |||
34 rGlnThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGln 50
      |||
168 ATTGTCTCTGCGACCGCGTAAACAGCTGACCTTATTTGAAAGAAAACA 217
      |||
51 lLeuAlleuTrpGlnProValThrAlaAspLeuIleValLysLysG 67
      |||
218 AGTCATTTTGTGTAATGCATCTGATGTGACAAATGGAAGCCCAT 267
      |||
67 nValHisPhePheValAsnAlaSerAspValAlaSerValLysAlaHisL 84
      |||
268 TAAATGTGACGGGAATTCATGACAGTGTCTTGTGACAGCGTGAAGAT 317
      |||
84 euAsnValSerGlyIleProCysSerValLeuAlaAspValGlnAsp 100
      |||
318 CTATTATCAACAGCAGATTTCACACAGACAGTACGCCCCGAGCCCTCCG 367
      |||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaSerAl 117
      |||
368 ATGCTACTATGACAGATATCATCTACATAAATGAAATCTATCTGTGATAG 417
      |||
117 aSerTyrTrpGlnGlnIleThrHisSerLeuAsnGlnIleTyrSerTrpIleG 134
      |||
418 AATTATATACAGAGCAGCATCTGATATGCTTACAAAATCCACATTTGA 467
      |||
134 lUpheIleThrGlnAlaGHisProAspMetLeuThrLysIleHisIleGly 150
      |||
468 TCCATATTGAGAAAGTACCCACCTATATGTTTAAAGTTTCTGAAAAGA 517
      |||
151 SerSerPheGlnLysTyrProLeuTyrValLeuLysValSerGlyLysG 167
      |||
518 ACAACAGCCAAAATGACCATATGATGATGACTGTGGAATCCATGCCAGAG 567
      |||
167 uGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184

```

```

568 AATGATCTCTCTGCTTCTGCTTGTGTCATAGGCCAT..... 608
184 IuPrIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
608 ..... 608
201 PheTrpGlyIleIleGlyGlnIYrThrAsnLeuLeuArgLeuValAspPh 217
608 ..... 608
217 eTyValMetProValValAsnValAspGlyTyTrpSerTrpIysl 234
609 ..AATGATGTGAGAGAAAGACCGTTCTTGTATGCGAAACATCATGTC 656
234 yAsnArTrpMetTrpArgIlyAsnArGserPheTyTrAlaAsnArHisCys 250
657 ATGGGAAGACCTGAATAGCACTTGTCTGCAACACAGTGCTAGGA 706
251 ILeGlyThrAspLeuAsnArGAsnPheAlaSerLysHisTrpCysGluG 267
707 AGCTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGACTTATCTTG 756
267 uGlyAlaSerSerSerSerCysSerGluTrpTrpTrpCysGlyLeuTrpPro 284
757 AGTCAGAACCAAGAGTGAAGGAGCGCTAGTTCTTGAAGAAATATC 806
284 IuSerIuProGluValIlyAlaValAlaLaserPheLeuArgAsnIle 300
807 AACCAATTAACATACATACATGATCATGATTCATCTCCAGCATATAGT 856
301 AAsnGlnIleLeuAlaTrpIleSerMetHisSerTrpSerGlnHisIleVa 317
857 GTTTCATATCTCTATACAGAAAGTAAAGCAAAACCATGAGAACTGT 906
317 IPrProTrpSerTrpTrpArgSerLysSerLysAspHisGluGluLeuS 334
907 CTTCTAGTACGAGTGAAGAGTCTGCTATGATTAACAACATAATAAT 956
334 eTyLeuValAlaSerGluAlaValAlaArgAlaIleGlyLysTrpSerLys 350
957 ACCAGTATACATGCGCATGCGTCAGAAACCTTATACCTAGCTCTG 1006
351 ThrArgTrpThrHisGlyHisGlySerGluTrpLeuTyIleuAlaProG 367
1007 AGCTGGGAGCATGATGATGATGATTTGGGCATCAATATTCCTT..... 1051
367 yGlyIlyAspArTrpIleTyTrpAspLeuGlyIleLysTySerPheThr 384
1052 .....TAC 1054
384 IeGlyLeuArgArTrpTrpGlyTrpPheLeuLeuProGluArgTyTr 400
1055 ATCAAAACCAACCTGTAGAGAACTTTGCCGCTGCTAAATAGCTTG 1104
401 ILeLysProThrCysArgGlyAlaPheAlaValSerLysIleLeuArg 417
1105 GCATGTCATTAGCAATGTT 1123
417 rHisValIleArgAsnVal 423
seq_name: sr_rodent:09JHN6
seq_documentation_block:
ID 09JHN6 PRELIMINARY; PRT; 422 AA.
AC 09JHN6;
DT 01-OCT-2000 (Tremblrel. 15, created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE CARBOXYPEPTIDASE R (THROMBIN-ACTIVATABLE FIBRINOLYSIS INHIBITOR)
DE (110032P04RIK PROTEIN).
DE CPB2 OR TAFI OR 110032P04RIK.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20341711; PubMed-10678383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA Campbell W., Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT whereas carboxypeptidase N is not.";
RL J. Immunol. 165:1053-1058(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-20201996; PubMed-10739389;
RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA Van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
RT inhibitor.";
RL Thromb. Haemost. 83:297-303(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE-21085660; PubMed-11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Koehliwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staudil F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzairelli J., Nomberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AB021968; BAB03402.1; -
DR EMBL: AF164524; AAF62385.1; -
DR EMBL: AK004045; BAB23141.1; -
DR HSSP: P00730; 2CTC.
DR MEROPS: M14.009; -.
DR MGP: MGI:1891837; Cpb2.
DR InterPro: IPR003146; Proper_M14.
DR InterPro: IPR000834; Zn_carboxept.
DR Pfam: PF02244; Proper_M14; 1.
DR Pfam: PF00246; Zn_carboxept; 1.
DR PRINTS: PR00765; CRBOXYPTSEA.
SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

```

alignment_scores:

```

Quality: 1592.50 Length: 422
Ratio: 4.576 Gaps: 3
Percent Similarity: 82.464 Percent Identity: 73.460

```

alignment_block:

```

US-09-980-881-1 x 09JHN6 ..
Align seg 1/1 to: 09JHN6 from: 1 to: 422

```

```

18 ATGAACTTTGCAAGCTTGCAGTCTTGTACCACTGTTCTCTGTGTA 67
||||||| :|||:||||| |||:|||||
1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTrp... 61 16
68 GCAGCATGTCCTGCGCTCCAGAGTGGCAAGTCCAGTCCAGTCTCTCCGA 117
||||||| ||||||| ||||||| ||||||| ||||||| |||||||
16 uGlnHisGlyPheAlaPheGlnSerIlyGlnValLeuSerAlaLeuProA 33

```

```

118 GAACCTTAGGCAAGTTCAAGTTTACAGAACTTACTACAACTATGAG 167
|||||
33 rgnrSerArgInValIglInleuInAsnLeuThrThrTyrCln 49
|||||
168 ATTGTCTCGGAGCGGCTAACAGCTGACCTTTGTGAGAAAGAAAACA 217
|||||
50 ValValLeuThrPglProValThrAlaIglInleuIglInLysLysG 66
|||||
218 AGTCCATTTTGTAAATGATCATGTGATGACATGACATGTAAGCCCAT 267
|||||
66 uValHisPhePheValAsnAlaSerAspValAlaSerValLysAlaH 83
|||||
268 TAAATGTAGCGGAATTCATGACAGTCTTGTGCGACAGCTGGAAGAT 317
|||||
83 euAsnValSerArgIleProPheAsnValIleuMetAsnValIglIn 99
|||||
318 CTATTTCACAGCAGATTTCCACAGACAGTACGCCCCGAGCTCCGC 367
|||||
100 LeuIleIglInIglInThrPheAsnAspThrValSerProAlaAlaSer 116
|||||
368 ATGCTACTATGACAGTATCATCTCATTAATGAATCTATTCTGATAG 417
|||||
116 aserTyrTyrGlnIglInThrHisSerLeuAsnGlnIleTyrSerTrp 133
|||||
418 AATTATTAAGTACAGGCACTCTGATATGCTTACAAAATCCACATTGA 467
|||||
133 luValIleThrIglInIglInHisProAspMetLeuGlnLysIleTyr 149
|||||
468 TCCCATTTGAGTACCCACCTATGTTTAAAGTTTGTGAGAAAAGA 517
|||||
150 SerSerPheGlnLysTyrProLeuTyrValLeuValSerGlnLysG 166
|||||
518 ACAAAACAGCCAAAATGCCATATGATGATGATGATGATGATGATG 567
|||||
166 uGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 183
|||||
568 AATGATCTCTCTGCTTCTGCTGTTGCTGATGATGATGATGATG 608
|||||
183 luTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199
|||||
608 ..... 608
200 PheHisGlyLysGlnAsnLeuTyrThrArgLeuLeuArgHisValAsp 216
|||||
608 ..... 608
216 eTyrIleMetProValMetAsnValAspGlyTyrAspTyrThrTrpLys 233
|||||
609 ..AATGCAATGTGGAGAAAGAACGTTCTTATGCGCAATCATTCAT 656
|||||
233 yAsnAlaArgMetTrpArgLysAsnArgSerAlaHisLysAsnAlaArg 249
|||||
657 ATGGAGACAGCCTGATAGCACTTGTCTCCAAACACCTGCTGAGAGA 706
|||||
250 ValGlyThrAspLeuAsnArgPheAlaSerLysHisTrpCysGlnLys 266
|||||
707 AGTGCATCTCAGTCTCATGCTCGAAACCTACTGTGAGCTTATCTGT 756
|||||
266 sGlyAlaSerSerSerSerGlnThrTyrCysGlyLeuTyrTrpG 283
|||||
757 AGTGAAGACAGAAAGTGAAGCAGTGGCTACTTCTTGAGAAAGAAAT 806
|||||
283 luSerGlnProGlnValLysAlaValAlaAspPheLeuArgArgAsn 299
|||||
807 AACGAGATTAAAGCATATCAGATCATCTCCGCAAGCTATATCTGT 856
|||||
300 AsnHisIleLysAlaTyrIleSerMetHisSerLysGlnGlnIleLe 316
|||||
857 GTTTCATATTCCTATACAGCAAGTAAAGCAAGACCATGAGAACTGT 906
|||||
316 upPheProTyrSerTyrAsnArgSerLysSerLysAspHisGlnGlnLe 333

```

```

907 CTCTAGTGCCAGTACAGCACTGCTGCTATGATGACAAAACATGTA 956
|||||
333 eLeuValAlaSerGlnIleValAlaValArgAlaIleGlnSerLysAsn 349
|||||
957 ACCAGTATACAGATGGCCATGGCTCAGAAACCTATACCTAGCTCCG 1006
|||||
350 ThrArgTyrThrHisGlySerGlySerGlnSerLeuTyrLeuAlaPro 366
|||||
1007 AGTGGGAGCAGATTGATCATGATTTGGCATCAAAATATCGTT..... 1051
|||||
366 yGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheH 383
|||||
1052 .....TAC 1054
383 leGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGlnArg 399
|||||
1055 ATCAAGCCACCTGTAGAGAGCTTTGGCCGCTGCTCAAAATAGCTGT 1104
|||||
400 IleLysProThrCysAlaGlnIleValAlaValAlaIleSerLysIleVal 416
|||||
1105 GCATGCTATTAGGAAT 1120
|||||
416 pHisValIleArgAsn 421

```

```

seq_name: sp_r0dent:O9QZF0
seq_documentation_block:
ID O9QZF0 PRELIMINARY; PRT; 422 AA.
AC O9QZF0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE U.
GN CPB2.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CHAPLJ;
RA He Y.C., Broze G.;
RT "Isolation and characterization of mouse liver carboxypeptidase B
RT gene."
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF186188; AAF00528.1; -.
DR HSSP; P00730; SCPA.
DR MEROPS; M14.009; -.
DR MGD; MGI:1891837; CpB2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxepr.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF02246; Zn_carboxepr; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SEQUENCE 422 AA; 48832 MW; B53FFB09943954E CRC64;

```

```

alignment_scores:
Quality: 1586.50 Length: 422
Ratio: 4.572 Gaps: 3
Percent Similarity: 82.227 Percent Identity: 73.223

```

```

alignment_block:
US-09-980-881-1 x O9QZF0 ..

```

```

Align seg 1/1 to: O9QZF0 from: 1 to: 422

```

```

18 ATGAAGCTTTCGAGCTTCGACCTCTTATGACCAATGTTCTCTGTGA 67
|||||
1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTyr...G 16
|||||
68 GCAGATGCTCTGCGGCTTCAGAGTGGCCAAATGTTTACGCTGCTTTCGA 117
|||||

```

```

16 uGlnHisGlyPheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProA 33
118 GAACCTTAGGCAAGTTCAGTTCATACAGAAATCTTACTACAAATATGAG 167
133 rGthSerArgGlnValGlnLeuGlnAsnLeuThrThrTrpGlyGlu 49
168 ATTGTTCTCTGCAGCCGGTAACAGCTGACCTTATTGTGAGAAAAAACA 217
50 ValValLeuTrpGlnProValThrAlaGlnPheIleGluLeuLysLysG1 66
218 AGTCATTTTTTTTGAATGCATGTGATGTCGACAAATGTGAAGCCCAT 267
66 ValAlaSerPheValAsnAlaSerAspValAspSerValLysAlaHisL 83
268 TAAATGTGACGCGAATTCGATGCACTGCTCTGCGACAGCTGGAAGAT 317
83 euAsnValSerArgTlleProPheAsnValLeuMetAsnAsnValGluAsp 99
318 CTTATTCAACAGCAGATTTCACAGACAGACAGTCCGCCCCGAGCTCCGC 367
100 LeuIleGlnGlnGlnThrPheAsnAspThrValSerProArgAlaSerAl 116
368 ATCTACTATGAAACAGTATCAGTACCTAAATGAATCTATCTTGGATAG 417
116 ASerTrpTrpGlnGlnIleTrpHisSerLeuAsnGlnIleTrpSerTrpIleG 133
418 AATTTATACAGAGGACATCCTGATATGCTTACAAAATCCACATTTGGA 467
133 lValLeuThrGlnGlnHisProAspMetLeuGlnLysIleTrpIleGly 149
468 TCCCTATTTGAGAAAGTACCCACTATATGTTTAAAGTTTCGGAATAA 517
150 SerSerPheGlnLysTrpLeuTrpValLeuLysValSerGlyLysG1 166
518 ACAACAGCCAAATAATGCCATATGATGACTGTGGAATTCATGCCAGAG 567
166 uGlnArgIleLysAsnAlaIleTrpIleAspGlyIleHisAlaArgG 183
568 AATGATCTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608
183 lUTrpIleSerProAlaPheCysLeuTrpPheIleGlyTrpValThrGln 199
608 ..... 608
200 PheHisGlyLysGlnAsnLeuTrpThrArgLeuAlaGlnHisValAspRh 216
608 ..... 608
216 eGlyTlleMetProValMetAsnValAspGlyTrpArgTrpTrpLysL 233
609 ..AATGCAATGTGAGAAAGAACCGTTCTTCTATGCGAACAATCATTTGC 656
233 yAsnArgMetTrpArgLysAsnAlaGSerAlaHisLysAsnAsnArgCys 249
657 ATCGGAACGACCTGAATAGCAACTTGTCTCCAAACACTGGTGTAAGA 706
250 ValGlyTrpAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGly 266
707 AGGTGCATCTCAGTTCCTCATGCTCGGAAACCTAAGTGCATTTATCTG 756
266 sGlyLysSerSerSerSerCysSerGlnThrTrpCysGlyLeuTrpProG 283
757 AGTCAGAACGACAGTGAAGGAGTGGCTAGTTCCTTGAAGAAATATTC 806
283 lUserGlnProGlnValLysAlaValAlaAspPheLeuArgTrpAsnIle 299
807 AACGAGATTAAAGCATACATCAGATGCAATTCATCTCCACAGATTAAT 856
300 AspHisIleLysAlaTrpIleSerMetHisSerTrpSerGlnGlnIleLe 316
857 GTTTCCATTTCTTATACAGCAAGTAAAGCAAGACATGAGGAGACTGT 906
316 uPheTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrp 333

```

```

907 CTCTAGTACGCAAGTTCAGAGTTCGTTGCTATTTGACAAAACCTAGTAAAT 956
333 erLeuValAlaSerGlnAlaValAlaArgAlaIleGlySerPheAsnLysAsn 349
957 ACCAGGTATACACATGGCCATGGCTCAGAAACCTTACTACTACTCTGCG 1006
350 ThrArgTrpThrHisGlySerGlySerGlySerLeuTrpLeuAlaProG1 366
1007 AGGTGGGACGATGTGATGATGATGATGATGATGATGATGATGATGATG 1051
366 yGlySerAspAspTrpIleTrpAspLeuGlyIleLysTrpSerPheThr 383
1052 .....TAC 1054
383 leGluLeuArgAspThrGlyArgTrpGlyPheLeuLeuProGlnArgTrp 399
1055 ATCAAAACCCACCTGAGAGAGCTTTGCGCGTCTGCTAAATACTGCTG 1104
400 lIeLysProThrCysAlaGlnAlaLeuAlaAlaIleSerLysIleValTr 416
1105 GCATGTCATTAGCAAT 1120
416 pHISValIleArgAsn 421

seq_name: sp_rodent:Q9EQV9

seq_documentation_block:
ID Q9EQV9 PRELIMINARY; PRT; 422 AA.
AC Q9EQV9:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PRE-PROCARBOXYPEPTIDASE R.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RX MEDLINE=20471387; PubMed=11021404;
RA Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA Okada N., Okada H.;
RT "Molecular cloning and partial characterization of rat
RT procarboxypeptidase R and carboxypeptidase N.";
RL Microbiol. Immunol. 44:719-728(2000).
DR EMBL: AB042598; BAB18617.1; -.
DR HSSP: P00730; 2CTC.
DR InterPro: IPR003146; Proper_M14.
DR InterPro: IPR00634; Zn_carboxept.
DR Pfam: PF02244; Proper_M14; 1.
DR Pfam: PF00246; Zn_carboxept; 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
KW Carboxypeptidase.
SQ
SEQUENCE 422 AA; 48826 MW; FFFD32A51A9366C8 CRC64;

alignment_scores:
Quality: 1568.50 Length: 422
Ratio: 4.507 Gaps: 3
Percent Similarity: 82.464 Percent Identity: 72.275

alignment_block:
US-09-980-881-1 x Q9EQV9 ..
Align seg 1/1 to: Q9EQV9 from: 1 to: 422

18 ATGACATTTGACAGCTTGACGCTTGTGACCATGTTCTCTCTGTGA 67
||||| :|||:||||| :|||:|||||
1 MetLysLeuTrpGlnGlnGlnValLeuValAlaIleLeuTrp...G1 16
68 GCAGCATGTCCTGCGTCCAGAGTGGCAAGTTCTAGCGCTCTGCTGA 117
||||| :|||:||||| :|||:|||||

```

```

16 ulysHISgLYleuAlaPheGlnSerGlyHisValLeuSerAlaLeuPro 33
118 GAACCTCTAGGCAAGTTCAGTCTCTACAGAACTTCTACTACACATATGAG 167
|||||
33 rghSerArGlnValGlnLeuLeuGlnAsnLeuThrThrTrpGlu 49
168 ATTTGTTCTGGCAGCCGGTACAGCTGACCTTATGTGTGAAGAAAAACA 217
|||||
50 ValValLeuTrpGlnProValThrAlaGluPheIleGluLysLysGlu 66
218 AGTCGATTTTTTTGTAATGATCATGTGTCGACAAATGTGAAGCCCAT 267
|||||
66 ValAlaSPhePheValAsnAlaSerAspValAsnSerValLysAlaTrp 83
268 TAAATGTGAGCGGAATTCATGCAAGTCTTCTTGGCAGACGTGAAGAT 317
|||||
83 euAsnAlaSerArGlieProPheAsnValLeuMetAsnValGluAsp 99
318 CTATTTCACAGCAGATTTCCACAGCAGCAGTCAAGCCCGACCTCCGC 367
|||||
100 LeuIleGlnGlnThrSerAsnAspThrValSerProArGlaSerSe 116
368 ATGCTACTATGACAGATTCATCTCAATAATGAATCTATTCTTGATG 417
|||||
116 rSerTrpTrpGlnGlnTrpHisSerLeuAsnGluIleTrpSerTrpIleG 133
418 AATTTCATGAGAGGATCTGATATGCTTACAAAATCCACATTTGA 467
|||||
133 luvAlaIleThrGlnGlnHisProAspMetLeuGlnLysIleTrpIleGly 149
468 TCCTCATTTGAGAGTACCCACTCTATGTTTAAAGTTTCTGGAAGA 517
|||||
150 SerSerTrpGluLysTrpProLeuTrpValLeuLysValSerGlyLysGlu 166
518 ACAAGACGCAAAATGCCATATGATGATGTCGTGGAAATCCATCCG 567
|||||
166 uHisAlaGValAlaAsnAlaIleThrPheAspCysGlyIleHisAlaArg 183
568 AATGATCTCTCCTGCTTCTGCTTGTGATTCATAGCCAT 608
|||||
183 lurPheSerProAlaPheCysLeuThrPheIleGlyTrpValThrGln 199
608 ..... 608
200 PheHisGlyLysGluAsnThrTrpThrArgLeuLeuArgHisValAsp 216
608 ..... 608
216 eTrpIleMetProValMetAsnValAspGlyTrpAspTrpThrTrpLys 233
609 ..AATGAAATGTGAGAAAGAACCGTTCTTCTATGCGAACAATCATATGC 656
|||||
233 yAsnAsnArgMetTrpArgLysAsnArgSerValHisMetAsnAsnArgCys 249
657 ATGGAAGACAGCTGAATAGCAACTTTGTCCAAACACTGCTGTGAGGA 706
|||||
250 ValGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGlu 266
707 AGGTGATCCAGTTCCTCATGCTCGAACAACCTACTGTGGACTTATGCT 756
|||||
266 sGlyAlaSerSerPheSerCysSerGluThrTrpCysGlyLeuTrpG 283
757 AGTCAGAACCAAGTGAAGAGCAGTGGTCTGTTCTTGAGAGAAATATC 806
|||||
283 lueSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299
807 AACCAAGTTTAAAGCATACATGACATGATCATCTCCACGATTAAGT 856
|||||
300 AsnHisIleLysAlaTrpIleSerMetHisSerTrpSerGlnIleLe 316
857 GTTTCATATTCCTATACAGAGTAAGTAAGCAAGACCATGAGGAATGT 906
|||||
316 upPheTrpTrpSerTrpAsnArgSerLysSerLysAspHisGlnGluLeu 333

```

```

907 CTCTAGTACGCAAGTGAAGCAGTTGCTGCTATGTGACAAAAGTAAAT 956
|||||
333 erLeuValAlaSerGluAlaValArgAlaIleGlnSerIleAsnLysAsn 349
957 ACCAGTATACACATGCGCATGCTGCGAGAAACCTTATACCTAGTCTCG 1006
|||||
350 ThrArgTrpThrHisGlySerGlySerGlnSerLeuTrpLeuAlaProG 366
1007 AGGTGGGAGCATGATGATCTATGATTTGGGATCAATATTCGTTTCAT 1055
|||||
366 yGlySerAspAspTrpIleTrpAspLeuGlyIleLysTrpSerPheThr 383
1056 C..... 1057
383 leGluLeuArgAspThrGlyArgTrpGlyPheLeuLeuProGluArgPhe 399
1058 ..AAGCCACCTGTAGAGAACTTTGCGGCTGCTCTTAAATAGCTTG 1104
|||||
400 lLeLysProThrCysAlaGluAlaLeuAlaValSerLysIleAlaTr 416
1105 GCATGCTATTAGGAAT 1120
|||||
416 PHisValIleArgAsn 421

```

seq_name: sp_human:Q9NTI8

seq_documentation_block: PRELIMINARY; PRT: 198 AA.

```

AC Q9NTI8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE BA139H14.2 (CARBOXYPEPTIDASE B2 (PLASMA)) (FRAGMENT).
GN CPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tromans A.;
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137141; CAB92622.1; -.
DR HSSP; P00730; ICPX.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF02446; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPRTASEA.
KW Carboxypeptidase.
FT NON_TER 198
FT SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;

```

alignment_scores:

quality: 1029.00 length: 197
ratio: 5.223 gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.492

alignment_block:

US-09-980-881-1 x Q9NTI8 ..

Align seg 1/1 to: Q9NTI8 from: 1 to: 198

```

18 ATGAAGCTTTGACAGCTTTGACAGTCTTGTACCAATGTTCTCTGTGCA 67
|||||
1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGlu 17
68 GCAGCATGTTCTGCGTTCCAGAGTGGCAAGTCTGAGTCTCTGCTTA 117
|||||
17 uGlnHisValaPheAlaPheGlnSerGlyGlnValLeuAlaIleLeuPro 34

```

```

118 GAACCTTAGCAAGTCAAGTCTACAGAACTTACTACACATATAG 167
|||||
34 rghrSerArgInValGlnValLeuGlnAsnLeuThrTrpTyrGlu 50
|||||
168 ATTGTTCTGCGACCGGTACAGCTGACCTATTGTGAAGAAAAACA 217
|||||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValIleValSylSylSyl 67
|||||
218 AGTCATTTTTTTTGTAAATGCATCTGATGTGACAAATGAAAGCCCAT 267
|||||
67 nvalHisPhePheValAsnAlaSerAspValAspAsnValIleAlaHisL 84
|||||
268 TAATATGACCGGAATTCATGAGAGTCTCTGCGACAGCTGGAAGAT 317
|||||
84 euAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
|||||
318 CTAAATTCACAGCAGATTTCACAGACACAGTACGCCCCGAGCTCCGC 367
|||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaSerAl 117
|||||
368 ATGCTACTATGAACAGTATCTACTAAATGAAATCTATTCTTGATAG 417
|||||
117 aSerTyrTyrGlnGlnIleThrHisSerLeuAsnGluIleTyrSerTrpIleG 134
|||||
418 AATTATTAATGAGCAGCATCCCTGATATGCTTACAAAATCCATTTGCA 467
|||||
134 IuhelIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
|||||
468 TCCTCATTTGAGAGTACCCACTGTATGTTTAAAGTTTCTGAAAAGA 517
|||||
151 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerIleLysG 167
|||||
518 ACAACAGCCAAAATGCCATATGATGATGACGTGGAATCCATGCCAGAG 567
|||||
167 uGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgS 184
|||||
568 AATGATCTCTCGCTTCTGCTTGTGTTGCTCATAGCCCAT 608
|||||
184 IuTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
|||||
seq_name: sp Vertebrate: 09PUF2

```

```

seq_documentation_block:
ID 09PUF2 PRELIMINARY: PRT: 416 AA.
AC 09PUF2:

```

```

DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE HOMOLOG.
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RA "Screening of Bothrops jararaca pancreas cDNA library.";
RT Submitted (Sep-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF190274; AF01344.1; -.
DR HSSP: P09955; INSA.
DR MEROPS: M14.003; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_cardopept.
DR Pfam: PF02244; Propep_M14.1.
DR Pfam: PF00246; Zn_cardopept.1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
SQ SEQUENCE 416 AA; 47723 MW; 4F99854DD72B7A7 CRC64;

```

alignment_scores:

```

Quality: 715.00 Length: 406
Ratio: 2.668 Gaps: 6
Percent Similarity: 66.010 Percent Identity: 37.192
alignment_block:
US-09-980-881-1 x 09PUF2 ..
Align seg 1/1 to: 09PUF2 from: 1 to: 416
60 TTTCTGAGCAGCATGCTTCCGCTCCAGACAGTCCCAATGTTACTGTC 109
|||||
14 PheAlaGluThrThrValHisArgPheAspGlyGluLysValTyrArgVal 30
|||||
110 TCTTCTAGAACCTTAGCAAGTCTCAAGTCTACAGAACTTACTACAA 159
|||||
30 IThrProArgAsnGluAspGluValTyrPheLeuAsnTyrLeuAlaAsnI 47
|||||
160 CATATGAGATTGTTCTCTGCGACCGGTACAGCTGACCTTATGTGAAG 209
|||||
47 LeValGlnValAspPheTrpArgProAspSerValGluLeuValIleSala 63
|||||
210 AAAAAACAGTCCATTTTGTAAATGCATCTGATGTGACAAATGTGA 259
|||||
64 GluMetThrValAspPheArgIleGluAlaAspArgCysSerIleValAl 80
|||||
260 AGCCCATTTAAATGTGAGCGGAATTCATGACAGTGTCTGTGCGAGAG 309
|||||
80 uSerIleLeuGlnSerGlyLeuAsnTyrGluIleLeuIleAspAsnL 97
|||||
310 TGGAAATCTTATTATCAACAGCAGATTCCACAGACAGCTACGCCCGCA 359
|||||
97 euGlnAlaValLeuAspArgGlnLeuAspAsnHisAla.....Arg 110
|||||
360 GCCTCCGACATCGTACTATGACAGTATCATCTACATAATGAATCTATTC 409
|||||
111 ThrAlaGlyTyrAsnTyrGluLysTyrAsnSerTrpGluLysIleAspAl 127
|||||
410 TTGGATAGAAATTTATACGTAGAGGAGCATCTGATATGCTTACAAAATCC 459
|||||
127 aTrpThrAlaAspIleAlaAsnGluAsnProSerLeuValSerArgLeuG 144
|||||
460 ACATGGATCCCATTTTGTAAATGCATCTGATGTGACAAATGTTTCT 509
|||||
144 IuIleGlyThrThrPheGlnGlyArgProMetProLeuLysVal... 159
|||||
510 GGAAAAAGAACAAACAGCCAAATAATGCCATATGATGATGATGATGCA 559
|||||
160 GlyLysProGlyValAsnLysLysAlaIlePheIleAspCysGlyPheH 176
|||||
560 TGGCAGAGATGATCTCTGCTGCTTCTGCTGTGTGTCATV..... 602
|||||
176 salArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluA 193
|||||
602 ..... 602
193 laValArgThrTyrGlyLysGluThrIleMetThrGlnLeuLeuAsnLys 209
|||||
602 ..... 602
210 LeuAspPheTyrIleLeuProValLeuAsnIleAspGlyTyrValTyrSe 226
|||||
603 ....GCCCATATTCAGATGTGAGAAAGAACCGTCTTTCATTCGCAACA 648
|||||
226 rTrpLysGlnSerArgMetTrpArgLysThrArgSerValAsnAlaGlyS 243
|||||
649 ATCATTCATTCGAGACAGCTGATAGCAACTGTGTCTGCTCAACACAGCG 698
|||||
243 eTrhCysIleGlyThrAspProAsnArgAsnPhe...AspAlaAlaTrp 258
|||||
699 TGTGAGGAAGTGCATCCAGTCTCATGCTCGAAGCTGGAACCTACTGTGACT 748
|||||
259 CysSerValGlyAlaSerArgAsnProCysSerGluThrTyrCysGlySe 275

```



```

||||| ..... ||||| .....
306 tserginmetleuleupherprotlytyrthrserlysleupropia 323
892 accatgaggaactgtctgtatgagcagatgaaacagcttgcgtattgac 941
323 snhlsiglunspreulalatsvalalalslglthrspvalleu... 338
942 aaacacgtaaaatcacagatatnacatggccatggctcagaaactt 991
339 serthargtyrteultharagtlyrtylletyrglyproillegluserthrii 355
992 ataccatgacctctggagtgaggcagatggatctatgatttgagcatca 1041
355 etyrproilserglyserleuaspripralatyraaspleuglylet 372
1042 aatattcgtttac..... 1054
372 ynhlstrhrhealaphegluleuargasplyslylspheglyphelu 388
1055 ..... atcaaacccactgtatgaaagctttgscgtgt 1089
389 leuproglyserargtylelyspthrctysargtltthmetleulava 405
1090 ccttaaaatagcttggaatgcttagtaagaat 1120
405 llyspheilealalystrylleulysnhs 415

seq_name: sp_human:Q96B08
seq_documentation_block:
ID Q96B08 PRELIMINARY; PRT: 417 AA.
AC Q96B08;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOHETICAL 47.4 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADRENAL CARCINOMA;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC015338; AAI15338.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 417 AA: 47367 MW; EBBB9B27F5D5AF9 CRC64;

alignment_scores:
Quality: 653.50 Length: 422
Ratio: 2.394 Gaps: 8
Percent Similarity: 64.692 Percent Identity: 33.649

alignment_block:
US-09-980-881-1 x Q96B08 ..

Align seg 1/1 to: Q96B08 from: 1 to: 417
24 CTTTGCAAGCTTGCAAGCTTGCAAGCTTGCTTGTGTGAGACGA 73
1 Metleulaleuleuvalleuvalthvalalaleulalaserlalnsh1 17
74 TGTCTTGGCG...TTCCAGAGTGGCAAGTTTAACTGCTCTCTTGA 120
17 sgllygllylnhlspheglulgllylvalrpheargvalasnvalglua 34
121 CTTTGAAGCAAGTTCAAGTTCTAACAAGTTTACTACACATAGAGATT 170
34 spgluaasnhs1leasn1lelleargtuleulalaserthtrhlnle 50
171 GTTCTGTGGAGCGGTAACAGTGTGACTTATGTGAAGAAAAAAGT 220

```

```

51 Asprhethrlyspkarservaltlnrglnleltsprohlsertnya 67
221 CCATTTTGTGTAATGCAATGATGTCGACAAATGGAAGGCCATTGA 270
67 lasprrhealvalylvalslglunspthrvalthrglualasnvalleu 84
271 ATGTGAGCGGAATTCATGCAAGTGTCTTGTGCGACAGCTGGAGATCTT 320
84 ysglmsnsluleuglnlyrlylvalleulleserlnleulargasnval 100
321 ATTCAACAGCAGATTTCACACAGACAGTCAGCCCCGAGCTCCGCAATC 370
101 valglualaglnrphaspservargval.....Arglatrhclgln 114
371 GTACTATGAACAGATCATCACTCAATGAATGAATCTATCTGTGATGAAT 420
114 ssetlyrtyllyltyrlyasnlystrpeltlnhrlleglualatprthgln 131
421 TTATAACTGAGAGCATCTGATATGCTTACAAAATTCACATGATGATCC 470
131 lnvalalatrhrgluaasproalaleulleserargservalileglythr 147
471 TCATTTGAAGAGTACCCACTCTATGTTTAAAGTTCTTGGAAGAAGACA 520
148 Thrpheluglyargalalietytleuleulysval...gllylsalagl 163
521 AACAGCCAAAATGCCATATGATGATGACTGTGGAATCCATGCCAGAAAT 570
163 ygluaasnlysprolalalrphemetaspcysglprhnsalalarglnt 180
571 GGATCTCTGCTGTCTGTGCTGTGCTGATCA..... 602
180 rplleserproalrphcysglnttrphevallarglualavalargthr 196
602 ..... 602
197 Tyrglylarginlleglnvalthrgluleuleuasplysleuasphe 213
603 .....GGCCAT.....A 609
213 rvalleuprovalleuasnleaspglyrtylletytrhtrpthrlyls 230
610 ATCGAATGTGGAAGAAAGACGCTTCTTATGCGAAACATTCATGATC 659
230 earlgrhethrprarglysthratserthnlsrthrglysercyslle 246
660 GGAACAGACCTGAATAGCAATTGTGTCCCAACACATGATGTGAGGAAG 709
247 Gltlytrpksnargasnprhe...Aspralaglytrpcysglulegl 262
710 TGCATTCAGTTCCTCATGCTCGGAAACCTTACTGTGACTTATTCGTGAT 759
262 yalaservargasnprocyasprgluthtyrcysgllyproalalalul 279
760 CAGAACCAAGATGAAGGAGCAGTGGCTAGTTCTTGAGAAGAAATATCAAC 809
279 erglulysgluthtyrlalaleulalaspheileargasnlylsleSer 295
810 CAGATTAAAGCATATCATCAGATGCATTCATCTCCAGCATATAGTGT 859
296 serlleulalalytleuthrlnhlssetyrsertlnmetkeltely 312
860 TCCATATTCCTATCAGAGTAAACCAAGACCAATGAGAGACTGCTGC 909
312 rprotysetyrtalalyrlylsleuglylualasnvalgluleuasn 329
910 TAGTAGCAGTAGAAGCAATTCGTGCTATGTGACAAACTAGTAATAATAC 959
329 lalaleulalysalatrhrvalylsleuleu...Alaservleuhslythr 344
960 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGGAG 1009

```



```

910 TAGTAGCAGTGAAGCAGTTCGCTCTATTGACAAACTAGTAAATACC 955
    :::::::::::::::::::::
329 snleuAlaValAlaValIleu...AlaThrIleuTyGlyThr 344
    :::::::::::::::::::::
960 AGGTATACATGATGGCTGCTGAGAAACCTTATACCTAGCTCTGGAGG 1009
    :::::::::::::::::::::
345 LysThrThrTyGlyProGlyAlaThrThrIleuProAlaIleuGly 361
    :::::::::::::::::::::
1010 TGGGAGCAGATGATCTATGATTTGGGCATCAATATTCGTTTACA 1055
    :::::::::::::::::::::
361 YserAspAspTrpAlaTyAspIleuGlyIleuTySerPheThr 376
seq_name: sp.rodent:p97597

```

```

seq_documentation_block:
ID P97597 PRELIMINARY; PRT; 412 AA.
AC P97597;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MAST CELL CARBOXYPEPTIDASE A PRECURSOR (FRAGMENT).
GN R-CPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE DAWLEY;
RX MEDLINE=97149430; PubMed=8996238;
RA Lutzelschab C., Pejler G., Aveskogh M., Hellman L.;
RT "Secretory granule proteases in rat mast cells. Cloning of 10
RT different serine proteases and a carboxypeptidase A from various rat
RT mast cell populations."
RL EMBL: U67914; AAB48267.1; -.
DR HSSP: P09955; INSA.
DR MEROPS: M14.010; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Signal; Carboxypeptidase.
FT NON_TER 1 1
FT SIGNAL 1 1
FT CHAIN 11 412 MAST CELL CARBOXYPEPTIDASE A.
SQ SEQUENCE 412 AA; 47944 MW; DAB59555FC49137D CRC64;

```

```

alignment_scores:
    Quality: 642.50      Length: 400
    Ratio: 2.530        Gaps: 7
    Percent Similarity: 63.500      Percent Identity: 35.750

```

alignment_block:

US-09-980-881-1 x P97597 ..

Align seg 1/1 to: P97597 from: 1 to: 412

```

84 TTCGAGATGGCCAGTCTTACGCTCTTCTAGAACTTAGGCAAGT 133
    :::::::::::::::::::::
16 PheAspArgGluValPheArgValIleuGlnAspGluIleuGlnAl 32
    :::::::::::::::::::::
134 TCAAGTCTACGAACTTACTACAAACATATGATGTTCTCTGGCAGC 183
    :::::::::::::::::::::
32 AserIleuLysAsnLeuThrGlnThrIleuLysPheTrpTyP 49
    :::::::::::::::::::::
184 CGGTAAAGCTGACCTTATTTGAAAGAAAACAGTCATTTTGTGTA 233
    :::::::::::::::::::::
49 roAspAlaIleuHisAspIleuAlaValAsnMetThrValAspPheArgVal 65

```

```

234 AATGCACTGATGTCGACAATGTGAAGCCCATTTAAATGAGCGGAT 283
    :::::::::::::::::::::
66 ThrGluLysGluSerGlnThrIleGlnSerThrLeuGlnIleHisLys 82
    :::::::::::::::::::::
284 TCCATGCAAGTGTCTGCTGGCAGACGTGGAAGATCTTATCAACAGCA 333
    :::::::::::::::::::::
82 LysPtyGluIleuIleuLeuAsnAspLeuGlnGluIleuAspLysGlnP 99
    :::::::::::::::::::::
334 TTTCAGACGACAGATGAGCCCGAGCCCGCATCGTCTATGAACAG 383
    :::::::::::::::::::::
99 he.....AspValLysGluGluIleuAlaGlyArgHisSerTyraIaLys 113
    :::::::::::::::::::::
384 TATCACTCACTAAATGAATCTTATCTTGATGAATCTTAACTGAGAG 433
    :::::::::::::::::::::
114 TyraAsnAspTrpAsnLysIleValSerTrpThrGluLysMetValGlu 130
    :::::::::::::::::::::
434 GCATCTGATATGCTTACAAAATCCACATGAGTCCTCATTTGAGAACT 483
    :::::::::::::::::::::
130 ShisProGluMetValSerArgIleuLysIleGlySerThrValGluAsp 147
    :::::::::::::::::::::
484 ACCCACTCTATGTTTAAAGTTCTTGGAAGAAGAACAGCCAAAT 533
    :::::::::::::::::::::
147 snProLeuTyraIleuLysIle...GlyArgLysAspGlyIleuArgLys 162
    :::::::::::::::::::::
534 GCCATATGATGATGATGATGATCCATGCCAGAGAAATGATCTCTCCGC 583
    :::::::::::::::::::::
163 AlaIlePheMetAspCysGlyIleHisAlaArgGluTrpAlaSerProAl 179
    :::::::::::::::::::::
584 TTTCCTGCTGTGTTTATA.....GCCCAATAT. 611
    :::::::::::::::::::::
179 aPheCysGlnTrpPheValTyGlnAlaIleLysSerTyGlyLysAsnL 196
    :::::::::::::::::::::
611 .....
196 YsIleMetThrLysLeuLeuAspArgMetAsnPheTyraIleuProVal 212
    :::::::::::::::::::::
612 .....
213 PheAsnValAspGlyTyriIleTrpSerTrpThrLysAspArgMetTrp 229
    :::::::::::::::::::::
623 AAGAACCCGCTCTTCTTACGCAACATATCTGACGCAACAGACCTGA 672
    :::::::::::::::::::::
229 GlyAsnAsnArgSerLysAsnProAsnSerThrCysIleGlyThrAspLeu 246
    :::::::::::::::::::::
673 ATAGCACTTTGTCCTCAAAACACTGGTGAGGAGGTGCATCCAGTCC 722
    :::::::::::::::::::::
246 snArgAsnPhe...AspValSerTrpAspSerSerProAsnThrAspAsn 261
    :::::::::::::::::::::
723 TCATGCTCGAAGACCTACTGTGACTTTATCTGATGCAACACAGAACT 772
    :::::::::::::::::::::
262 ProCysLeuSerValTyraGlyProAlaProGluSerGluLysGluThr 278
    :::::::::::::::::::::
773 GAAGCAGAGCGCTAGTTCTTGAGAAGAAATATCAACCGATTAAGCAT 822
    :::::::::::::::::::::
278 rLysAlaValAlaThrAsnPheIleArgSerHisLysAsnSerIleuLysAla 295
    :::::::::::::::::::::
823 ACATGCACTGATCATCTACCTCCACATATAGTGTTCATATTCCTAT 872
    :::::::::::::::::::::
295 YrIleThrPheHisSerTyriSerGlnMetLeuLeuPheProTyriGlyTy 311
    :::::::::::::::::::::
873 ACAGCAAGTAAAGCAAGACCATGAGAACTGTCTCTATGACCGCATGA 922
    :::::::::::::::::::::
312 ThrIleLysLeuProProAsnHisGlnAspLeuLysValAlaArgIle 328
    :::::::::::::::::::::
923 AGCAGTCTGCTATGACAAACTATGTAATAATACAGGTATPACATG 972
    :::::::::::::::::::::
328 eAlaThrAspValLeu...SerSerArgTyriGluThrArgTyriIleTyG 344
    :::::::::::::::::::::
973 GCCATGCTGCAAGAAACCTTATACCTAGCTCTGAGAGGTGGGAGCATGG 1022
    :::::::::::::::::::::
344 TyProIleAlaSerThrIleTyriLysThrSerGlySerSerLeuAspTrp 360

```

```

1023 ATCTATGATTTGGGCATCAATATTCGTTAC..... 1054
1055 .....ATCAACCCACCTGTA 1070
377 sglySerglyPheLeuLeuProGlySerArgIleLeuProThrCysL 394
1071 GAGAGCTTTTCCGCTCTCTAAATPAGCTTGGCATGTCATTAGCAAT 1120
394 ysgIuThrPheValArgGluAlaValArgThrTyrLeuGlnIuLeuHis 410

seq_name: sp_rudent:Q9CVD1

seq_documentation_block:
ID Q9CVD1 PRELIMINARY; PRT; 279 AA.
AC Q9CVD1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 2210008M23RIK PROTEIN (FRAGMENT).
GN 2210008M23RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57Bl/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawal J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Sakai K., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderrelli R., Barsch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RA *Functional annotation of a full-length mouse cDNA collection.*;
RT Nature 409:685-690(2001).
RL EMBL; AK008678; BAB25826.1; -.
DR HSSP; P09955; INSA.
DR MGD; MGI:1923953; 2210008M23RIK.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF00246; Zn_carboxypept. 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; UNKNOWN_1.
DR NON_TER 279
SQ SEQUENCE 279 AA; 31970 MW; 29468AFA0B67FOAA CRC64;

alignment_scores:
Quality: 531.00 Length: 226
Ratio: 3.493 Gaps: 4
Percent Similarity: 67.257 Percent Identity: 45.575

alignment_block:
US-09-980-881-1 x Q9CVD1 ..
Align seg 1/1 to: Q9CVD1 from: 1 to: 279
489 CTCATGTTTAAAGGTTCTGAAAGACAAACAGCCAAATATGCGAT 538
:::|||||::: |||||::: || |||:::

```

```

1 MetYrValLeuLysIle...GlyLysAspArgProAsnLysProAlaPh 16
539 ATGATTTAGCTGTGGATTCATGCCAGAGAAATGATCTCTCTGCTTCT 588
:::|||||:::|||||:::|||||:::|||||:::
16 ePheLLeAspCysGlyPheHisAlaArgLutPrLLeSerProAlaPheC 33
569 GCTTGTGTTCAATA..... 602
33 ysgIuThrPheValArgGluAlaValArgThrTyrLeuGlnIuLeuHis 49
602 ..... 602
50 MetYsArgLeuLeuAspGluLeuAspPheTyrValLeuProValAs 66
603 .....GCCATPATGCAATGTGAGAAAGA 627
66 nileAspGlyTyrValTyrThrThrPalLysAspArgMetThrPargLysT 83
628 ACCGTTCTTTCTATGCAACAATCATTCGATCGGAACAGACCTGAAATAC 677
:::|||||:::|||||:::|||||:::
83 hrArgSerThrThrAlaGlySerSerCysPheGlyValAspProAsnArg 99
678 AACCTTGTCTCCAAACACAGCTGGTGAGAGAGTGATGATCAGTTCTCATAG 727
|||||:::|||||:::|||||:::|||||:::
100 AsnPhe...AspAlaGlyTyrPcysGluValGlyAlaSerArgSerProCy 115
728 CTCGGAACCTACTGTGACTTTATCTGAGTACAGACAGACAGAGAGAG 777
:::|||||:::|||||:::|||||:::|||||:::
115 sSerAspThrTyrCysGlyProThrProGlySerLysGluThrLysA 132
778 CAGTGGCTATGTTCTTGAGAGAAATATCACAGATTAAAGCATATCATC 827
|||||:::|||||:::|||||:::|||||:::|||||:::
132 lAlaLeuAlaAspPheLeuArgGlnAsnLeuSerSerIleLysAlaTyrLeu 148
828 AGCAGCATATCATCTCCAGCATATGAGTTTCATATTCCTATACACG 877
:::|||||:::|||||:::|||||:::|||||:::
149 ThrValHisSerTyrSerGlnMetLeuTyrProTyrSerTyrAspTy 165
878 AAGTAAAGCAAGACATGAGAACTGCTCTAGTACGACAGTGAAGAG 927
:::|||||:::|||||:::|||||:::|||||:::
165 rLysLeuProGluAsnTyrGlnGluLeuAsnAlaLeuValLysGlyAla 182
928 TTCGCTGATTTGACAAACTAGTAAATACAGATATACATGCGCAT 977
:::|||||:::|||||:::|||||:::|||||:::
182 lAlaYsGluLeu...SerThrLeuHisGlyThrLysTyrThrTyrGlyPro 197
978 GCGTCAGAAACCTTATACCTAGCTCTGAGGTGGGAGATTCGATGATCA 1027
|||||:::|||||:::|||||:::|||||:::|||||:::
198 GlyAlaThrThrIleTyrProAlaAlaGlyLysSerAspAspThrPalTy 214
1028 TCATTTGGGCATCAATATTCGTTTACA 1055
||| |||||:::|||||:::|||||:::
214 rAspGlnGlyIleLysTyrSerPheThr 223

seq_name: sp_mammal:O9YV85

seq_documentation_block:
ID O9YV85 PRELIMINARY; PRT; 419 AA.
AC O9YV85;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE AI (BC 3.4.17.1).
GN PCPAl.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99192816; PubMed=10092856;
RA Darnis S., Juge N., Marino C., Aviles F.X., Pulgserver A., Chalik J.C.,
RA Guo X.J.;

```

"Cloning, sequencing and functional expression of a cDNA encoding RT porcine pancreatic preprocarboxypeptidase A1.";
 RL Eur. J. Biochem. 259:719-725(1999).
 DR EMBL: AF076222; AAD17690.1; -
 DR HSSP: P09954; 1PCA.
 DR MEROPS: M14.001; -
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxypept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxypept. 1.
 DR PRINTS: PR00765; CRBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolase; Carboxypeptidase.
 SQ SEQUENCE 419 AA; 47235 MW; 84B4CB57B714FC1 CRC64;

alignment_scores:

Quality: 516.50 Length: 391
 Ratio: 2.207 Gaps: 12
 Percent Similarity: 59.847 Percent Identity: 33.760

alignment_block:

US-980-881-1 x 09TV85

Align seg 1/1 to: 09TV85 from: 1 to: 419

```

24 CTTTGACGCTTGACGCTTGATCCCATTTGCTCTTGATGACGCA 73
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
4 LeuLeuIlePheSerValLeuLeuGlyGlyValLeuAlaLysGluAsp 19
74 TGTCTCCGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 123
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
20 .....PheValGlyHisGlyValLeuArgGlyLeuSerValAspArg 33
124 STAGGCAAGTTCAAGTTCTACAGATTTCTACACATATGAGATGTT 173
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
33 LuAlaGluValGluLysValLysGluLeuGluLysLeuGluHisLysGlu 49
174 CTC.....TGCGAG...CCGGTAAACGCTGACCTTATGTAAGAAA 214
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
50 LeuAspPheThrPargGlyProAlaArgProGlyPheProIleAspVal 66
215 ACAAGTCATTTTGTAAATGACATGTCATGTCGACCAATGTGAAGCCC 264
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
66 9.....ValProPheProSerIleGluAlaValLysValP 78
265 ATTAAATGATGACGCAATTCATGCAAGTCTTGCTGCGACAGCTGA 314
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
78 heLeuGluAlaHisGlyIleArgGlyThrIleMetIleGluAspValGlu 94
315 GATCTATTT.....CAACGACAGATTTCCAGCAGACAGCTGACGCC 355
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
95 LeuLeuLeuAspArgIleGluGluGluMetPheAlaSerGluIleArgAl 111
356 CCGAGCCCTCCGATCTGAC...TATGACAGATTCATCTCACTAAAGAAA 402
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
111 ArgGlyThrIleThrPheAsnGlyAlaThrGlyHisThrIleGluGluI 128
403 TCTATCTTGATGATGATTAATTAACGACGACGATCTGATATGCTTACA 452
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
128 IeArgPhePheMetAspIleLeuValAlaGluHisProGluLeuValSer 144
453 AAAATCCATGATGATCCCATTTTGAGAGTACCCCATGTTTAA 502
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
145 LysLeuGluHisIleLysSerGlyGluIleArgProIleArgValLeuL 161
503 GGTTCCTGGAAGAAAGAACACCCAAAAT.....GCATATGGA 543
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
161 sPheSer.....ThrGlyLysHisAsnArgProAlaIleIleTrp 174
544 TTGACTGTGGAATCCATGCCAGAGAAATGATCTCTGCTTGTGCTTG 593
   ||| :|||:|||||:|||||:|||||:|||||:|||||:

```

```

174 IeAspThrGlyIleHisSerArgGluTrpValThrGluAlaSerGlyVal 190
594 TGGTTCATTA..... 602
191 TrpPheAlaLysLysIleThrGluAspTrpGlyGluAspProAlaPheTh 207
602 ..... 602
207 AlaIleLeuAspAsnLeuAspIlePheLeuGluIleValThrAsnPro 224
603 .....GCCCATATCGAATGTGGAGAAAGACCGT 632
224 spGlyPheAlaPheThrHisSerGluAsnArgMetTrpArgLysThrArg 240
633 TCTTTATGAGAACATCATTCATGCAAGCAAGACCTGATTAAGCAACT 682
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
241 SerArgThrSerGlySerPheCysValGlyValAspProAsnArgAsnTr 257
683 TGTCTCCAAACACTGATGTGAGGAGGATCCCATCTCCATGCTCGG 732
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
257 P...AspAlaGlyPheGlyGlyAlaGlyAlaSerSerAsnProCysSer 273
733 AAACCTACTGTGACTTATCTGAGTACAGACCAAGTGAAGGCAAGTG 782
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
273 LuThrTrpHisGlyLysPheProAsnSerGluValGluValLysSerIle 289
783 GCTAGTTCTTGAGAAAGAAATATCAACGATTAAGCATATCAATGCAAT 832
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
290 ValAspPheValAsnAspHisGlyAsn...IleLysAlaPheIleSerI 305
833 GCATTCATCTCCGACATTAAGTGTTCATATTCCTATACAGAAAGTA 882
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
305 eHisSerTrpSerGluLeuLeuLeuTrpProGlyGlyTrpLysThrGlu 932
883 AAAGCAAGCAATGAGAACTCTCTAGTACCAAGTGAAGTTCGT 932
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
322 IeProAlaAspLysAspGluLeuAspGluIleSerLysSerAlaValAla 938
933 GCTATTGACAAACTAGTAAATATACAGGATATACATGATGCGATGCTC 982
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
339 AlaLeu...ThrSerLeuTrpGlyThrLysPheGluTrpGlySerIle 354
983 AGAAACCTTATACCTAGCTCTGAGCTGGAGCTGGAGACGATGATATGAT 1032
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
354 eThrTrpIleTrpGluAlaSerGlyGlyThrIleAspTrpThrLysAsn 371
1033 TGGCATCAAAATATGCTTTTACA 1055
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
371 IeGlyIleLysTrpSerPheSer 378

```

seq_name: sp_human:096QN3

seq_documentation_block:

ID 096QN3 PRELIMINARY; PRT; 417 AA.
 AC 096QN3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE CARBOXYPEPTIDASE A2 (PANCREATIC).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=PANCREAS;
 RA Strausberg R.;
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC007009; AAD07009.1; -
 SQ SEQUENCE 417 AA; 46756 MW; D18B6784DFA862 CRC64;

alignment_scores:

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2002, 15:18:17 ; Search time 2123.6 Seconds

(without alignments)
15500.773 Million cell updates/sec

Title: US-09-980-881-1

Perfect score: 1573
Sequence: 1 agaaaattctgttggatg.....aaaaaaaaaaaaaaaaaa 1573

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mn:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match Length	DB ID	Description

1	1528	97.1	1546	9	AB011969	AB011969 Homo sapi
2	1372	87.2	1728	9	HUMPCPBX	M75106 Human prepr
3	1372	87.2	1749	6	I16100	I16100 Sequence 2
4	1372	87.2	1749	6	I33526	I33526 Sequence 2
5	1309	83.2	1715	9	BC007057	BC007057 Homo sapi
6	918	58.4	1272	6	AR086324	AR086324 Sequence
7	656	41.7	1421	10	AB021968	AB021968 Mus muscu
8	656	41.7	1490	10	AF164524	AF164524 Mus muscu
9	654	41.6	1430	10	AF186188	AF186188 Mus muscu
10	652.6	41.5	1425	10	AB042598	AB042598 Rattus no
11	475.4	30.2	58097	9	AL157758	AL157758 Human DNA
12	409	26.0	437	11	G05971	G05971 human STS W
13	173.6	11.0	181	11	G59992	G59992 SHGC-130613
14	128	8.1	180680	9	AL137141	AL137141 Human DNA
15	124	7.9	927	5	AR072911	AR072911 Sequence
16	123.8	7.9	1327	5	AF190274	AF190274 Bothrops
17	123.4	7.8	921	6	I67699	I67699 Sequence 3
18	123.4	7.8	1215	6	I67698	I67698 Sequence 1
19	123.4	7.8	1251	4	SSC13375	AJ133755 Sus scrof
20	123.2	7.8	451	11	G53326	G53326 SHGC-82320
21	120.8	7.7	1251	4	DOGZAP47	D78348 Dog mRNA fo
22	109	6.9	1263	6	AB7530	AB7530 Sequence 11
23	107.6	6.8	1622	9	HUMCARWC	M27717 Human mast
24	107.6	6.8	1622	11	G28614	G28614 human STS S
25	107.6	6.8	1674	9	BC012613	BC012613 Homo sapi
26	107.4	6.8	999	6	AS1908	AS1908 Sequence 72
27	107.4	6.8	999	6	AR085871	AR085871 Sequence
28	107.4	6.8	1053	6	AS1904	AS1904 Sequence 68
29	107.4	6.8	1053	6	AR085868	AR085868 Sequence
30	107.4	6.8	1263	6	AS1896	AS1896 Sequence 60
31	107.4	6.8	1263	6	AR085861	AR085861 Sequence
32	107.4	6.8	1284	6	AS1913	AS1913 Sequence 77
33	107.4	6.8	1284	6	AR085875	AR085875 Sequence
34	107.4	6.8	1310	9	HSR224866	AJ224866 Homo sapi
35	107.4	6.8	1389	9	BC015104	BC015104 Homo sapi
36	107.4	6.8	1454	9	BC015338	BC015338 Homo sapi
37	104.2	6.6	1053	6	AX000417	AX000417 Sequence
38	104.2	6.6	1053	6	AX000418	AX000418 Sequence
39	104.2	6.6	1059	6	AS1917	AS1917 Sequence 81
40	104.2	6.6	1059	6	AS1921	AS1921 Sequence 85
41	104.2	6.6	1059	6	AR085878	AR085878 Sequence
42	104.2	6.6	1059	6	AR085881	AR085881 Sequence
43	102.8	6.5	927	9	S40234	S40234 mast cell c
44	101	6.4	1870	6	A67356	A67356 Sequence 11
45	101	6.4	1870	6	A87561	A87561 Sequence 42

ALIGNMENTS

RESULT	1	1546 bp	mRNA	linear	PRI 02-FEB-2000
AB011969	AB011969	Homo sapiens mRNA for carboxypeptidase B-like protein, complete cds.			
LOCUS	AB011969				
DEFINITION	AB011969				
ACCESSION	AB011969				
VERSION	AB011969.1	GI:6855463			
KEYWORDS	carboxypeptidase B-like protein.				
SOURCE	Homo sapiens	CDNA to mRNA.			
ORGANISM	Homo sapiens				
REFERENCE	Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.				
AUTHORS	1 (sites)				
TITLE	Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase B from human plasma				
JOURNAL	J. Biol. Chem. 266 (32), 21833-21838 (1991)				
REFERENCE	92042093				
AUTHORS	2 (sites)				
TITLE	Matsumoto, A.				
JOURNAL	Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase B from human plasma				
REFERENCE	Unpublished (1998)				

REFERENCE 3 (bases 1 to 1546)
AUTHORS Matsumoto, A.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1998) Akira Matsumoto, Kobe University School of Medicine, Department of Radiation Biophysics & Genetics; Kusunoki-cho 7-5-1, Chuo-Ku, Kobe, Hyogo 650-0017, Japan (E-mail: amatemed.kobe-u.ac.jp, Tel: 81-78341-7451)
COMMENT Sequence updated (10-Apr-1998).
FEATURES
source
1..1546
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
20..1102
/note="carboxypeptidase B-like protein in human brain"
/codon_start=-1
/product="carboxypeptidase B-like protein"
/protein_id="BAA90475.1"
/db_xref="GI:6855464"
CDS
1546
polyA_site
/note="21 A nucleotides"
BASE COUNT 453 a 323 c 297 g 473 t
ORIGIN
Query Match 97.18; Score 1528; DB 9; Length 1546;
Best Local Similarly 99.48; Pred. NO. 0;
Matches 1534; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
1 agaaattgctgctggaatgaagcttgagcctgagctgctgtaaccattgctcct 60
3 AGAAAAATTCGTTGGATGAAGCTTTGACACCTTGACGCTCTGTACCCATGTTCTCT 62
61 tctgtgagcagcagctgctgctgctgctgctgctgctgctgctgctgctgct 120
63 TCTGTGAGCAGCAGCTGTGCGCTTTCAGACTGGCAGAGTCTAGCTGCTGCTTCTTAA 122
121 cctctagagcaagctcgaagcttctacagaactcttactacaacatgagatgctctc 180
123 CCTCTAGGCAAGTTCAGATCTTACAGATCTTACTACAACTATGAGATTTCTCTGGC 182
181 agccggtacacagctgaccttctgtagaagaaaaaacaagctcatttttctgtaagcat 240
183 AGCCGGTACAGCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTCTTAAATGAT 242
241 ctgagtcgacaaatgtagaagcccatctaaatgtagcagaaatcctcagtgctctgc 300
243 CTGATGTCGACAAATGTGAAGCCCATTTAAATGTGAGGGAATTCATGCACTGCTTGC 302
301 tggagagcgtgtagaagctcttctacagaagatctcacaagcagctcagcccgag 360
303 TGGGAGATGTGGAGATCTTATTCACAGCAGATTTCCACAGACAGTCAAGCCCGCAG 362
361 cctcgcagatgtagtaagacagtaactcaactaaatgtaattattcttgtagaagat 420
363 CCTCCGATCTACTATGAAAGATACATCCTCACTAAATGAATATTTCTTGGATAGAT 422
421 ttaatactgagagcagctcctgatalgcttaacaaatccacatgagctcctcattgaga 480
423 TTATAACTGAGAGCAGCTCCTGATATGCTTACAAAAATCCACATTTGATCCCATTTGAGA 482
481 agtacccactctatgctttaaaggcttctggaagaagaacaacagcaaaaaatgcatat 540
483 AGTACCACACTATGTTTAAAGGTTCTGGAAGAAAGAACAGCAAAATAATGCCATAT 542
541 ggattgctgtggaatcagcagagaatgagatcctcctgcttctgctgtggttca 600

543 GGAATGACTGTGGAAATTCATCCAGAGATGAGATCTCTGCTTTCCTGTTGCTTCA 602
601 taagccatcatcgaatgtagaagaagacgcttcttctatgcaaacatcatgacatcgc 660
603 TAGGCCATTAATCGAATGTGGAGAAAGAACGCTTCTTTATGCGAACATATCTGATC 662
661 gaacagactgaatagcaacttgcctcacaacactggtgtagaagaagtgatcagatt 720
663 GAACAGACCTTAATGAGAACTTGTGCTTCAAAACACTGCTGTGAGGAAGTGCATCAAGT 722
721 cctatgctcgaaacacactgtagcttctcctgagtcgaaccgaatgtagaagag 780
723 CCTATGCTCGAAACCTTACTGTGACTTTTATCCGATCGAAGACCAAGTGAAGCGAG 782
781 tgcctagcttctgagaagaataatcaaccagatgaagcaatacagcagcatgcat 840
783 TGGCTAGTTCTTGGAGAAATATCAACGATTAAGCATATCATCGATGATTCAT 842
841 accccagcatatagtgcttccatattcctatcaacgaagtaaaagaacacatgag 900
843 ACTCCACAGCATATGAGTGTTCATATTCCTATACAGCAAGTAAAGCAAGACCATGAG 902
901 aactgctcgttagagcagtagaagcaagtcgctgctatgtaaaaactagtaaaatacca 960
903 AACTGCTCTAGTAGCCAGTGAAGCAGTTCGCTATTGAAGAAATTAAGTAAATACCA 962
961 ggtatcacatgagcagctgctcagaacacatcctatcctagcctcctgaggtgtagaagcatt 1020
963 GGTATACAGCATGAGCCATGCGCAGAAACCTATATCTATACCTGCTCGAGAGTGGACGATT 1022
1021 ggatctatgatttgggacatcaaatattcgtttacatcaaacacacactgtagaagcatt 1080
1023 GGACTATGATTTGGGCTATCAAAATATTCGTTATCATCAAAACCCAGCTGATGAGAAGCTTT 1082
1081 tgcgcgtctctcctaagaatgcttgagcagctatgaagatgtaaaagccctgatttta 1140
1083 TGCGGCTGCTCTAAAGAGCTTGCGCATGTATTAAGCAATTTAAAGCCCTGATTTTA 1142
1141 tcaattgctcgcgtatcttlaattactgattcagcaagaacccaatcatgtaacgat 1200
1143 TCATTCGCTCGCTGATTTTAATTTAGTATTCAGATTCAGCAAGCAAAATCATGTATCAGAT 1202
1201 taattttaaagttatccgtatgatttgataaagaatttccattccttgctgctcga 1260
1203 TATTTTAAAGTTATTCGCTGATTTGATTAAGAAATTTCCATTCCTGCTGCTGCTG 1262
1261 agaaccttaataagctgaccttgcacatgaagcagagctcagctgcttcttaccct 1320
1263 AGAACCTTAATAGTGTCTATTTGCCATTTAAGGACAGACTAGGGTTCATGCTTTTACCTT 1322
1321 ttaaaaaaaaattglaaaagctagttactacttcttcttgaatttctgagcgttgaat 1380
1323 TTAATAAAAAAATGTAAAGCTAGTTACCTATTTTCTTGTGATTTTTCGACGTTTGAAT 1382
1381 agccatctcaagcaacttgcagcttgcagctagcactccatcaagaagtttaataaagat 1440
1383 AGCCATCTCAAGCAACTTTCACGCTTGGACATGACATCTCAAGCAAGTTTAAATCAAGAT 1442
1441 catctcagctgtagcatgtagtgcctactcaacaaagaagaggtggtcagaagaactta 1500
1443 CATCTCAGCTGATCATTTGATCTTACTCAACAAAGGAAGGAGTGTGAGAAATGATTA 1502
1501 aagattctgctccaatttcaataaattctctctctcctt 1544
1503 AAGATTCTGCTCCAATTTTCAATAAATTTCTTCTCTCTT 1546
RESULT 2
HUMCPBX HUMCPBX 1728 bp. mRNA linear PRI 07-JAN-1995
LOCUS Human prepro-plasma carboxypeptidase B mRNA, complete cds.
DEFINITION
ACCESSION M75106
VERSION M75106.1 GI:189686

KEYWORDS plasma carboxypeptidase.
SOURCE Homo sapiens liver cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1728)
AUTHORS Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
TITLE Isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma
JOURNAL J. Biol. Chem. 266 (32), 21833-21838 (1991)
MEDLINE 92042093
FEATURES
source location/Qualifiers
1..1728 /organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
1..1728 /gene="PCPB"
/gene="PCPB"
20..1291 /note="prepro-plasma carboxypeptidase B"
/codon_start=1 /protein_id="AA60042.1"
/db_xref="GI:189687"
/translation="MKLCSTLAVLPYVIFCEQHYFAFGSGVLAALPRTSRQVYLQN
LTTVEIVLMDPTADLIVKQYHEFVNASDVDAVKALHNSGIPCSVLADVEDLI
QOOISNDIVSPRASAYEDYHNEIYSWLEFTEHPDMITKIHGSSPEKPLVY
LVKSGKEDYKNAIWMIDGIDHAREMISPAFLMFIQHTQFGLIQTNLRLVDY
LMPVNDYGDYSWKNNMNRKNSFYNNHCIGDILNRNASKMCEGASSCSE
TYGLYPESEPEVKAVASFLNRNINOTIKAYISMSISQHIYEPVPSYTSKSKDHELS
LVASENRAIRKTSKNRTYTHGSEETIYLAPEGGDDMIYDLGISTYIELRDTGTY
GFLPERITKTKREAFPAVSKIAMHVIRN"
mat_peptide /gene="PCPB"
86..1288 /product="plasma carboxypeptidase B"
BASE COUNT 518 a 354 c 338 g 518 t
ORIGIN

Query Match 87.2% Score 1372; DB 9; Length 1728;
Best Local Similarity 90.3% Pred. No. 4.3e-307;
Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;
QY 1 agaaaatgctgtgtggatgaagcttgcagccttgacgtctgtacccattgtctct 60
DB 3 AGAAAATTGCTGTGTGGATGAAGCTTTGGACGCTTGACGTCCTTGACCAATGGTCTCT 62
QY 61 tctgtgaagcagatgtcttcgcttcagagtgagcagaattttagctgtctctcagaa 120
DB 63 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGCCCAAGTTCTAGCTGCTCTCTAGAA 122
QY 121 cctctagagcaagttcaagttctacagaactctactacaacatagatgtgtctctgc 180
DB 123 CCTTAGGCAAGTTCAAGTTCTACAGAACTTACTACAACATATGATGTGTTCTGTGCG 182
QY 181 agcgggtacagctgacattatgtgaaagaaaaacaagtcacatttttctgtaaatgcat 240
DB 183 AGCCCGTAAACACTGACCTTATTTGGAAGAAAAACAAGTCCATTTTGTGTAATGCAT 242
QY 241 ctgattgtcacaatgtgaaagcccatltaaatgtgtagcggaatttcacatgtctcttc 300
DB 243 CTGATGTGCACATATGAAAGCCCATTTAAATGTAGCGGAATTCATCAAGTGTCTTGC 302
QY 301 tggcagagctgtgaaagatcttactacaacagcagatttccaagcagacagcagcccgag 360
DB 303 TGGCAGAGGTGGAAGATCTTATTCAACAGCAGATTTCGAACAGACAGACCCCGAG 362
QY 361 cctcgcatctgactatgaacagatcactcactaaatgaatctatcttctgtagaagt 420
DB 363 CCTCGCATCTGACTATGAAACAGTATCACTCACTAAATGAATCTATCTTGATGAAT 422
QY 421 ttatacgaagagcattctgtatgcttacaataatcacattgattctcatttgaaga 480

DB 423 TTATTAAGTAGAGGACATCCTGATATGCTTACAAAAATCCACATTTGATCCTATTGGAGA 482
QY 481 agtaccactctatgttttaaaagttctctgaaagaacaaagcagcaaaatgcatat 540
DB 483 AGTACCCACTCATATGTTTAAAGTTTGTGGAAGAAACAAACAGCCAAATATGCATAT 542
QY 541 ggaattactgtggaatccatgccagaagaatgagatctctctctctctctctgtgtgttca 600
DB 543 GGATTGACGTGGAAATCCATGACAGAGAAATGATCTCTCTGCTTCTGCTGTGTCA 602
QY 601 taggcact----- 608
DB 603 TAGGCCATATTAACATATTCATGGATTAATAGGCCAATATACCAATCTCTGAGGCTTG 662
QY 609 -----a 609
DB 663 TGGATTTCATGTTATGCGGGTGTTAATGTGACGGTTATGACTCTCATGGAAGAAAAGA 722
QY 610 atcgaatgtgagagaagaacggtctctctatgacgaacatcatgcatcggaagacc 669
DB 723 ATCGAATGTGAGAAAGAACCCGTTCTTCTGATGCGAACAATCATGTCAGAACAGACC 782
QY 670 tgaatgcaacttgcctcacaacactggtgtgagagaagtgatctcagttctctatgct 729
DB 783 TGAATAGCACTTGTGCTCCAAACACTGGATGAGGAGAGGTGATCCAGTCTCATGCT 842
QY 730 cggaaacctactgtgacttatacctcgaatcagaacagaagtgaaggcagtgactagt 789
DB 843 CGGAACCTACAGTGTGACCTTATCCGATCAGAACACAGAGGAGGAGCGTGCATGTT 902
QY 790 tcttgaagaagaatcacaacagaltaaagacatcacatcgatgcatctactccagc 849
DB 903 TCTTGAGAAAGAAATATCAACACAGATTAAGCATACATCAGCATGCTATCTCCACAG 962
QY 850 atatagtgttccatattctctatcacagaagttaaagaacagacatgaggagactcttc 909
DB 963 ATATAGTGTTCATATTCCTTATACAGAAATGAAGCAAGACCTGAGAACTGTCTC 1022
QY 910 tagtaagcagtgaaagcagtcgtgctatgacaaactagtaaaaaatatacaggtatacac 969
DB 1023 TAGTACCAGTGAAGCAGTTCGTGCTATGTAGAGAAACTGTAAATATACAGGTATACAC 1082
QY 970 atggcctatgctcagaacattatcctatgctctcggaggtgaggagattgagatcatg 1029
DB 1083 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCTCGAGGTGGAGCGATTGGATCTAG 1142
QY 1030 atttggcctcaaatatgt----- 1050
DB 1143 ATTTGGCATCAAAATATGTGTTTACAATTGAACTTGAGATACGGGCACATACGATTTCT 1202
QY 1051 -----ttacatcaaacaccactgttagagaagcttttgcgctgtctcaaaa 1097
DB 1203 TGTGTCGGAGAGGTTTACATCAAAACCCAGCTGTAGAGAACTTTTGGCGGTCTCTTAAAA 1262
QY 1098 tagcttgcatgtcattaggaatgtttaaagccctgatttatacatctcgtccgtat 1157
DB 1263 TAGCTTGCGATGTCAATAGGAATGTTAATGCCCTGATTTTATCATCTGCTCCGTAAT 1322
QY 1158 tttaatctcattctcagaagaacaaatcatatgataatattttaaagtttatac 1217
DB 1323 TTTAATTTACTGATTCAGAGAACCAAAATCATGTATCATGATTAATTTTAACTTTATTC 1382
QY 1218 cgtagtttgaataaagatttctcattctcctgtgtctctgcagagaacctataaagtgtc 1277
DB 1383 CGTAGTTTGTATPAAAGATTTTCTATTCCTTGTTCTGTGCANAGAACCTAATAAGTGTCT 1442
QY 1278 acttggccatlaagcagacataggttcaatgtcttttacccttataaaaaaaattgttaa 1337
DB 1443 ACTTGCCTATTAAGCAGACATAGGGTTCAATGCTTTTACCCCTTAAAAAAAATTGTAA 1502
QY 1338 aagcttaagtactacttttcttcttgaatttgcagatttgactagcactcaagaagat 1397
DB 1503 AAGCTTAGTACCTACTTCTTCTTGTGATTTTGCAGCTTTGACTAGCCATCTCAAGCAACT 1562

QY 1398 ttcgacgtttgactagccatctcaagcaagtttaactaaagatcctcaagctgatat 1457
Db 1553 ttcgacgtttgactagccatctcaagcaagtttaactaaagatcctcaagctgatat 1622
QY 1458 tggatccctcaacaaagaagaagtggtcagaagatcatatgaagtttctgctccaa 1517
Db 1623 tggatccctcaacaaagaagaagtggtcagaagatcatatgaagtttctgctccaa 1682
QY 1518 tttcaataaattctctctctctcttaaaaaaaaaaaaaaaaaa 1563
Db 1683 tttcaataaattctctctctctcttaaaaaaaaaaaaaaaaaa 1728
RESULT 3
LOCUS 116100 1749 bp DNA linear PAT 03-APR-1996
DEFINITION Sequence 2 from patent US 5474901.
ACCESSION 116100
VERSION 116100.1 GI:1251008
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1749)
AUTHORS Drayna, D. T. and Ealon, D. L.
TITLE Antibodies to human carboxypeptidase B and methods of use thereof
JOURNAL Patent: US 5474901-A 2 12-DEC-1995;
FEATURES
source 1. 1749
location/Qualifiers
BASE COUNT 521 a 361 c 342 g 525 t
ORIGIN

Query Match 87.2%; Score 1372; DB 6; Length 1749;
Best Local Similarity 90.3%; Pred. No. 4.3e-307;
Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

QY 1 agaaaaatgctgttggtggaagagctttgagccttgagctctgttaccatgttctc 60
Db 24 AGAAAAATGCTGTGGGTGAAGCTTTCAGCCCTTGACGCTCTGTACCCATGTGTCCT 83
QY 61 tctgtgagcaagcagctctcgtctccagagtggtgcaagttctagctgctctctctgaa 120
Db 84 TCTGTGAGCAGACAGTCTCTCCGCTTTCAGAGTGCACCAAGTTCTAGCTGCTCTCTGA 143
QY 121 cctctaggaagttcaagttctacagaattcttactacaacatagagattgtctctg 180
Db 144 CCTCTAGGCAAGTTCAGATTCTACAGAACTTACTACAACTATGAGATTGTCTCTG 203
QY 181 agccggtaaagcgtgacattcttctgaaagaaaaaagaagtcattcttctgttaagt 240
Db 204 AGCCGGTAAAGCTGACCTTATTTGTGAAGAAAAAACAAGTCCATTCTTTTGTAAAT 263
QY 241 ctgagtcgacaatgtgaaagccatttaaatgtgagcgaattccatgagtgctctgc 300
Db 264 CTGATGTGCAACAAGTGAAGCCATTAAATGTGACGGGATTCATGCAAGTCTTGGC 323
QY 301 tggcagaagctggaaagatctattcaacagagattccaaagacagtcagcccccag 360
Db 324 TGGGAGAGCTGGAAGATCTTATTCACAGCAGATTTCACAGACAGAGTGAAGCCCG 383
QY 361 cctccgagatgctactaagaacagtatcaactcaactaaatgaattatcttctgagtag 420
Db 384 CCTCCGATGTACTATGAACAGATACACTCACTAAATGAATCTATCTTGGATGAAT 443
QY 421 ttaatactgagagcattcgtatgtcttaaaaaatccacattgagttccatctttaga 480
Db 444 TTATTAATGAGAGCACTCCTGATATGCTTTACAAAAATCCACATTGGATCCATTTAGA 503
QY 481 agtaccactctatgttttaaagttcttgaaagaacaaacagcaaaatgcatat 540
|||||

Db 504 AGTACCACCTGTANGTTTTAAAGTTTCTGGAAAAAACAACCAACCAAAATGCCATAT 563
QY 541 ggaattgactgtgaaatccatgcccagagaatgatctctcctgcttctgtgtgtca 600
Db 564 GGATTGACTGTGGAATCCATGCGACAGAAATGATCTCTCTGCTTTCTGCTTGCTGCA 623
QY 601 taagcaat----- 608
Db 624 TAGGCATATTAATCAATTTATGAGATATATAGGCATATACCAATCTCCTGAGGCTTG 683
QY 609 -----a 609
Db 684 TGGATTTTATGTTATGCCGGTGTTAAVTGGACGGTTATGACTACTCATGAAAAAGA 743
QY 610 atcgaatgtgagaagaagaccgttcttctctatgcaacaatcatctgcatggaacagacc 669
Db 744 ATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAACATATTCATTCGGAACAGACACC 803
QY 670 tgaatagcaacttgctccaaacactggtgtgaggaagtgcataccagttcctcatgct 729
Db 804 TGAATAGGACCTTGTCTCCAAACACTGCTGTGAGGAAGTGCATCCAGTTCCTCATGCT 863
QY 730 cggaaacctactatggaacttatctcgtatgagagcaagaagtgaagcagtgagtgatc 789
Db 864 CGGAACCTACTGTGACTTTATCTGTGAGTGAACCGAAGTAAAGCACTGAGCTAGTT 923
QY 790 tcttgagaagaataatcaacagatlaaagacatacatgaatcatcatcaccagc 849
Db 924 TCTTGAGAAAGAAATATCAACAGATTAAAGCATATCATGAGATCATATACCTCCACAGC 983
QY 850 atataagtttccatattctctatacagaagtaagaacaaagacatgaggaactgtctc 909
Db 984 ATATAGTGTTCATATTCCTATACAGCAAGTAAAGCAAGCAATGAGGAACCTGCTC 1043
QY 910 tagtaagcagtaagcagttcgtctatcttgacaaactagtaaaataaacaagtaatacac 969
Db 1044 TAGTAGCCAGAGAAAGCACTGCTGCTATTGAGAAACTAGTAAATAATCCAGGATATCAC 1103
QY 970 atggcattgctcagaacaccttatacctagctcctctggaagttggtggaagatgatact 1029
Db 1104 ATGGCCATGCGTCAGAAACCTTATACCTAGCTCCTGAGGAGGAGATGGATGTATG 1163
QY 1030 attgggcatcaaatatgct----- 1050
Db 1164 ATTTGGGATCAAAATTCCTGTTTACAAATGCAATTCAGATACGGGCATACGGATTC 1223
QY 1051 -----tcaacaacaaccacccctgtagaagaagctttgcccgtctctaaaa 1097
Db 1224 TGCATGCCGAGCGTTATCATCAAAACCCACCTGTAGAGAAAGCTTTGCCGCTGCTTAAA 1283
QY 1098 taagctggcattgcatatgaagaatttaaatgccccctgatttcaatctgctcgtat 1157
Db 1284 TAGCTTGGAATGTCATATGGAATGTAAATGCCCTGATTTTATCATTCGTGCTCGTAT 1343
QY 1158 ttaattactgattccagaagaagcaaatcatctgtatcaagattatcttaagttatc 1217
Db 1344 TTATTAATTAAGTATCCAGCAAGACCAAAATCAATGTATGAGATTATTTAAAGTTTATC 1403
QY 1218 cgtagtttgataaaagaatttccatctctgtgtctgtgcaagaagcaacaaatgaatg 1277
Db 1404 CGTAGTTTATATAAATATTTCTCTATCTTGTTCTGTGCAAGAACTTAATAATGCT 1463
QY 1278 acttgcattaaaggagactaaggtcatgctcttttaaccttaaaaaaaatgttaa 1337
Db 1464 ACTTTGCCATTAAGGCGAGACTAGGTTTCATGCTTTTATACCTTTAAAAAAATTTG 1523
QY 1338 aagcttagtactactcttcttcttgatcttgagcgtttgactaagcaatcaagaact 1397
Db 1524 AAGCTTAGTTACTACTTTTCTTTGATTTTTCAGACGTTTGAACATTCACATCAAGCAACT 1583
QY 1398 ttcgacgtttgactagccatctcaagcaagtttaactaaagatcctcaagctgatat 1457
Db 1584 TTGAGGTTGACTAGCACTCAAGCAAGTTTAATCAAAAGATCATCTACGCTGATCAT 1643
|||||

Oy 1518 ttccaataaattctctctccttaaaaaaaaaaaaaaaaaa 1563
 Db 1704 TTTCAATTAATTTCTCTTCCTTTAAAAA 1749
 RESULT 5
 BC007057 1715 bp mRNA linear PRI 12-JUL-2001
 LOCUS Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
 DEFINITION IMAGE:3934520, mRNA, complete cds.
 ACCESSION BC007057
 VERSION BC007057.1 GI:13937896
 KEYWORDS MGC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1715)
 AUTHORS Strausberg, R.
 JOURNAL Direct Submission
 Submitted (30-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: CLONTECH
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 16 Row: F Column: 21
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4503004.
 FEATURES
 source
 location/Qualifiers
 1..1715
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:12495 IMAGE:3934520"
 /tissue_type="Skeletal Muscle"
 /clone_lib="NIH-MGC_81"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 25..1296
 /codon_start=-1
 /product="carboxypeptidase B2 (plasma)"
 /protein_id="AA07057.1"
 /db_xref="GI:13937897"
 /translation="MKLCSLAVLPVIVLCEQHVFAFQSGOVLALPRTSROYVLON
 LMTTVEIVLWQPTADLVKQVHFFVNSVDVNAKHLNVGICSVLLAIVEDLI
 QOOISNDVSPRASASIEQYHSLNEYSIEFTTHERPMLKIHGSEFEKPYLV
 LKYSGEQAKNAIWDICGTHAREWISPAFLCFLHITQFVGLIOGYTLALNDFY
 TVPVVVDGVDYQWKKRNRKRSFYANNHCIGTDLNRRFAKNNCEBASSSCSE
 TVCGLVPESEPKAVASTLRNINIKALITSHASTSOHIVFYSTRSKSDHEELS
 LVASEAVRAIEKTSKNTRYTHGSEFLVILAPGGGDWVIYDLGIKYSFTLELDGTLY
 GFLPRLRYKTPCREAFVAVSKIAWHVIRNV"
 BASE COUNT 518 a 348 c 341 g 508 t
 ORIGIN
 Query Match 83.2%; Score 1309; DB 9; Length 1715;
 Best Local Similarity 89.4%; Pired. No. 1.6e-292;
 Matches 1512; Conservative 0; Mismatches 15; Indels 164; Gaps 3;

Oy 1 agaaattgctgttggaatgaagcttgcagccttgacgttcctgttaccattgttctc 60
 Db 8 AGAAATGCTGCTTGGATGAAGCTTTCAGCTTGACGCTCTGTATCCATGTGTCCT 67
 Oy 61 tctgtgaagcaagatgtcttcgcgttcacagatgggccaagctctagctcctcctagaa 120
 Db 68 TCTGTGACAGCATGTCTCGGCTTTCAGAGTGGCCAAAGTTCTAGCTCTTCTCTAGAA 127
 Oy 121 cctctgaagcaagttcaaatctctacagaatctctactacaacatgagattgtcttgc 180
 Db 128 CCTTAGGCAAGTTCAAAGTCTACAGAACTTACTACAACTATGACAGATGCTCTGCGC 187
 Oy 181 agccgttaacagcttaccttattgtgaagaaaaaagaagctctttttttaaagcat 240
 Db 188 AGCCGGTAACAGCTGACTTATTGTGAGAAACAAAGTCAATTTTGTAAATGAT 247
 Oy 241 ctgattgcacaatgtgaagcccaattaaatgtgagcggaaattcactgacgtgtctgc 300
 Db 248 CTGATGTCACAAATGTGAAGCCCATTTAAATGTAGCGGAATTCATGACAGTGTCTGC 307
 Oy 301 tggcagagtggaagatctttaaagcagatcttccaagacacagtcagcccgag 360
 Db 308 TGGCAGAGTGGAAGATCTTATTCACAGCAGATTTCCAAAGACACAGTCAGCCCGAG 367
 Oy 361 cctccgacatcgtactatgaacagatcactcactaaatgaatctatctgtgataagat 420
 Db 368 CCTCCGACATGCTACTATGAAACAGATACATCCTCAATGAAATGATTTCTGGATTAAT 427
 Oy 421 ttataactgagagatcctcgtatgtcttacaataatccacattgatacctcattgaga 480
 Db 428 TTATTAACGAGAGGACATCTCATATGCTTACAAAACCAATGTGATCCCATTTAGA 487
 Oy 481 agtaccacatcattgtttaaggtctctgaaagaaagaaagccaaatgagcatat 540
 Db 488 AGTACCCACTTATGTTTAAAGGTTTCTGGAAGAAAGCAAGCCAAATAATCCATAT 547
 Oy 541 ggaattgactgtgaatccatgcagagaatgagatctcctgcttctgtgtgttca 600
 Db 548 GGATTTGACTGTGAATCATCATGCCAGAGATGATCTCGTTCGTGTGTTCA 607
 Oy 601 taggcact----- 608
 Db 608 TAGGCCATTAATCAATTTCTATGGGAATAGGGCAATATACCAATCTCTGAGGCTTG 667
 Oy 609 -----a 609
 Db 668 TCGATTTCTATGTTATGCCGCTGTTATGTGATGATGTTATGACTCATGCAAAAAGA 727
 Oy 610 atcgaatgtggaagaagaacggtctctctatgcaacaatcattgcatcgaagagacc 669
 Db 728 ATCGAATGTGGAAGAAAGACCGTCTTCTATGGAACATCATTTGATCGGAACAGACC 787
 Oy 670 tgaatgaacaactgtctccaacaacactggtgtgagaaggtgcatcgaattctcattgct 729
 Db 788 TGAATGGAACCTTGTCTTCCAAACACTGGTGTGAGGAAGTGATCCAGTCTCTCATGCT 847
 Oy 730 cggaaacctactgtgaacttalcctcctagtcagaaaccaagaagtgaagcagtgactgt 789
 Db 848 CGGAACCTTACTGTGACTTTATTCCTAGTCAGAACCAAGAAAGTGAAGCGTGGCTAGTT 907
 Oy 790 tcttgagaagaataatacaaccagaattaaagcatatcagcatgcatctactccagc 849
 Db 908 TCTTGAAGAAATATCAACACAGATTAACCATATCAATCATGATCATCTATCTCCAGC 967
 Oy 850 atataggttccatctctctatcaacgaagtaaaagcaagccatgggaactgtctc 909
 Db 968 ATATAGTCTTCCATATTTCTATACAGAAAGTAAAGCAAGCAATGAGAACTGTCTC 1027
 Oy 910 tagtaagcagtgaaagcagttcgtgtctattgacaaactagtaaaatacaggtatcacac 969
 Db 1028 TAGAGGCACTGAAGCAGGTTGCTGCTATATGAGAAACTGTAATAAATACAGGTATACAC 1087

OY 970 atggccatggctcaaaacctatatactagctcctggagggtggagcattgattctatg 1029
|||||
Db 1088 ATGGCCATGGCTCAGAAACCTTATACGTCTGGAGGTGGGACGATGGATCTATG 1147
OY 1030 atttggatcaaatcatctcgt----- 1050
|||||
Db 1148 ATTTGGGATCAAAATTTCTGTTTACATTTGAACTTCGAGATACGGGACATACGATTTCT 1207
OY 1051 -----ttacatcaaacccaacctgtagaagaactttgcgcgtgtctctaaaa 1097
|||||
Db 1208 TGCTGCCGAGCGTTACATCAAAACCACCTGTAGAGAACTTTTGGCGTCTCTAAAA 1267
OY 1098 tagcttgcatgctcattgggaatgtttaatgcacctgatttatactcttgcctat 1157
Db 1268 TAGCTTGGCATGTCAATAGGAATGTTAAVGGCCCTGATTTTATCATTCCTGCTTCCGTAT 1327
OY 1158 ttaattctactgattcccaagaacaaatcatgtatcagaattatttaagtcttacc 1217
|||||
Db 1328 TTTAATTTACTGATTCACGACCAACCAATCATGTGATACGATTTATTTAAGTTTATC 1387
OY 1218 cgtagcttgcataaagaatttccctacttccctgtgtctgctgcagagaacctaaagtgc 1277
Db 1388 CGTAGTTTGTATTAAGATTTTCTATCTCTGTTCTGTCTCAGAGAACTTAATAAGTGTCT 1447
OY 1278 acttggccattaaggcagactagggctcagtctcttcttacccttcaaaaaaatgttaa 1337
Db 1448 ACTTTTGCATTAAGGACAGCTVAGGGTTCATGCTTTTAAACCTTTT-AAAAAAATTTGTA 1506
OY 1338 aagctagcttaacctacttcttcttgcatttgcagcttgcagctcacaagaact 1397
Db 1507 AAGTCTAGTACCTTCTTTTCTTTGATTTTGCAGCTTTGATAGCATCTCAAGCAACT 1566
OY 1398 ttgcagcttgcagctcactcctcaagaagtttaactcaagaatcatctcagctgatact 1457
Db 1567 TTGCAGCTTTGACCTACCTCAAGCAAGTTTGATTCATGATCATCTCAGCGTGATCAT 1626
OY 1458 tggatccctactcaaaaaagaagggtgtgtcagaagtaactaaagaattctgcctcaaa 1517
Db 1627 TGGATCTTACTCAACAAAAAGGAGGTGGTCAAGATACATTAAAGATTTCGTCCATATA 1686
OY 1518 ttccaataaa 1528
|||||
Db 1687 AAAAAAAAAA 1697
RESULT 6
AR086324 1272 bp DNA linear PAT 07-SEP-2000
LOCUS AR086324
DEFINITION Sequence 1 from patent US 5985562.
ACCESSION AR086324
VERSION AR086324.1 GI:10013090
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Morser,M,John and Nagashima,M.
TITLE Method of detecting thrombotic disease risk associated with plasma
carboxypeptidase B polymorphisms
JOURNAL Patent: US 5985562-A 1 16-NOV-1999;
FEATURES
Location/Qualifiers
source 1..1272
BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN
Query Match 58.4%; Score 918; DB 6; Length 1272;
Best Local Similarity 86.8%; Pred. No. 4.7e-202;
Matches 1104; Conservative 0; Mismatches 5; Indels 163; Gaps 2;
OY 18 atgaagcttggagccttcagctccttgcacattgtctctctgtgtagaagcattgc 77
|||||

Db 1 ATGAACCTTTGCAGCCTTGACGCTTGTATCCCATTTGTTCTCTCTGTGACAGCATGTC 60
OY 78 ttcgcgttccaaagatggtccaaagttcctagctgtctcttccctagaaccttagcagaattcaa 137
|||||
Db 61 TTCCGCTTTACAGGTGGCCCAAGTTCAGTGTGCTTCCTGTAAGAACCTTATGGCAAGTTCAA 120
OY 138 gtctcagaatcttactacaacatatgagatgttctctgtgcagccggttaacagctgac 197
|||||
Db 121 GTTCTCAACAATCTTCTCAACATATGAGATTGTTCTCTGGAGCGGCTAACAGCTGAC 180
OY 198 cttaattgtaagaaaaaacaagctcaatttttggtaaatgcatctgattgcacaattg 257
|||||
Db 181 CTTATTTGAGAAAAAACAAGTCCATTTTGTAAATGCACTGATGTGCACAATGATG 240
OY 258 aaagccatttaaatgtgcaggaattccaatgcagtgatctgtctgcgcagacgtgagaat 317
Db 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTCTGTGTCGACAGCGTGAAGAT 300
OY 318 ctatctcaacagaagatttccaagacacagtlcagcccccagcctcgcactgactat 377
Db 301 CTTATTTCAACAGCAGATTTTCCAAACGACAGTCAGCCGCCGACCTCCCATCGTACTAT 360
OY 378 gaacagtatcactcactaaatgaatcctatcttgcagtagaatttaaatgtagagcat 437
Db 361 GAAAGATATCACTCACTAATGAAATCTATCTTGGATGAAATTTAATCTGAGAGGCAT 420
OY 438 cctgatatgtctcaaaaaatccacatttgatccctcatcttgagaagtaccacactatgt 497
Db 421 CCGATATGCTTACAAAATTCACATTTGATGATCCTCATTTTGAGAGATACCACTCTATGTT 480
OY 498 ttaaaagttctggaagaagaacaacagccaaataatgcatatgattgactgtgaaac 557
Db 481 TTTAAAGTTTCTGGAATAAACAACAACAACCAAAAAAGCCATATGATGACTGTGGAATC 540
OY 558 catggcaagaatgagttctctcctgtcttctgtctgtgtgtaataagcat----- 608
Db 541 CATGCCAGAAATGAAATGATCTCTCTGCTTCTGTGTGTTCTATGAGCCATATTACTCAA 600
OY 609 ----- 608
Db 601 TTCTATGGGTAATAGGCAATATACCAATCTCCAGGCTTGATTTCTATGTATG 660
OY 609 -----aatcgaatgtgagaag 626
Db 661 CCGGTGTTAATGTGACGGTATGACTACTCATGGAAGAAAGATGATGAGAGAAAG 720
OY 627 aacggtcttctatgctgaagaacatcatctgcatgcagacagcactgatatgaactgtgc 686
Db 721 AACGTTCTTTCTATGCAAACTATTCATGCAAGCTGGAACAGACTGAAATGGAACCTTGTCT 780
OY 687 tccaacactgtgtgtaggaaggtgcatccagttccctcatgtctcggaaaacctactgtga 746
Db 781 TCCAAACTGTGTGAGGAGAGTGCATCCAGTTCCTCATGCTCGGAAACCTCTGTGGA 840
OY 747 cttaatcctgagtcagaacagaagtgaagcagtggtcagtttcttgcagaagaatlatc 806
Db 841 CTTTATCTTGAGTCAGAACACAGCAAGTGAAGGCGTGGTAGTTCTTGTGAGAAAGAAATATC 900
OY 807 aaccgaattaaagcatcaatcagcatgcatctatctatctccagataatagtgttccatat 866
Db 901 AACCAATTAAGCATATCATGCAATGATCATCTTACCTCCAGCATATAGTGTTCATATAT 960
OY 867 tccatacacaagaagttaaagaacacatgagaactctctctagtagcagtgaaagca 926
Db 961 TCCATATACAGGAAGTAAAGCAAGACATGAGAGACTGTCTCTATAGCATGCACTGAAGCA 1020
OY 927 gtctgtctatttgacaacaaactagtaaaataccaggatatacacatggtccattgcaaga 986
Db 1021 GTTCTGCTATTGAGAAACTAGTAAATAATACAGGTATACACATGCGCATGGCTCGAGAA 1080
OY 987 acctatacctagctctcctggaggtgggagcattgatactatgatttggcacaatat 1046
Db 1081 ACCTTATACCTAGCTCTCTGAGGTGGGACGATTGGAATCTATGATTTGGGACATCAATAT 1140

```

OY 1047 tegt-----ttac 1054
    |||
Db 1141 TCGTTTACATGAACTTCGAGATACGGGACATACGATTCCTTGTCGCCGAGGCTTAC 1200

OY 1055 atcaaacaccactgtgaagaagcttttgcgcgtctcttaaaatagcttgacatgcatc 1114
    |||
Db 1201 ATCAAAACCCACCTGTAGAGAACTTTTGCCTGCTCTTAAATAGCTTGCGATGTCATT 1260

OY 1115 aggaatglttaa 1126
    |||
Db 1261 AGGATGTTTAA 1272

RESULT 7
AB021968 1421 bp mRNA linear ROD 27-JUL-2000
LOCUS Mus musculus mRNA for carboxypeptidase R, complete cds.
DEFINITION AB021968
ACCESSION AB021968.1 GI:9558447
VERSION carboxypeptidase R.
KEYWORDS carboxypeptidase R.
SOURCE Mus musculus CDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Sato, T., Miwa, T., Akatsu, H., Matsukawa, N., Obata, K., Okada, N.,
Campbell, W. and Okada, H.
PRO-carboxypeptidase R is an acute phase protein in the mouse,
whereas carboxypeptidase N is not
J. Immunol. 165 (2), 1053-1058 (2000)
20341711
2 (bases 1 to 1421)
Sato, T.
Direct Submission
Submitted (28-DEC-1998) Tomoo Sato, Fukushima Hospital, Chouju
Medical Institute, 19-14, Azayamanka, Noyori-cho, Toyohashi, Aichi
441-8124, Japan (E-mail: tomoos@cp-ip.or.jp, Tel: 81-532-46-7511,
Fax: 81-532-46-4899)
FEATURES
SOURCE Location/Qualifiers
1..1421
/organism="Mus musculus"
/db_xref="taxon:10090"
17..1285
/codon_start=1
/product="carboxypeptidase R"
/protein_id="BAB03402.1"
/db_xref="GI:9558448"
/translation="MKLHGILIVAILYBOHPAROSGOVLSAPRTSNQLOLNL
TTTTVEVLMQPVTAEPTEKKKVEYFYNADVDVSAHNLVSRPVLNNVEDLIE
QQTENQVSPRASASTEYQHSLEITSFLEVTEDHPDLQKITYGSSEKYPVLY
KVSQKEQRIKNAIMIDCGIHAREMISPAFLWFIQVYTOHGENLTRLRLRHDFYI
MPVMNDGYDYTKKNMMNRKNSAHKNCVGTDLNRNRSKHKMCKGASSSSCSST
YCCGLYSESEVKAADVADFLRRNIDHIKAYISMSYSQQLFPYSYNSKSKDHEEELI
VASEAVATIESINKNTRYTHSGSESLYLA PGSDDMITVDLGIKYSTITLDRNTRGRG
FLPERIKPTCAEALAAIKIYMHVIRNT"
1421
polyA_site
BASE COUNT 418 a 320 c 305 g 378 t
ORIGIN
Query Match 41.7%; Score 656; DB 10; Length 1421;
Best Local Similarity 72.2%; Pred. No. 1.9e-141;
Matches 1009; Conservative 0; Mismatches 215; Indels 174; Gaps 5;

```

```

OY 124 ctaggcaagttcaagttcttaccagaactcttactacaacatatgagattgtctctgagc 183
    |||
Db 120 CCAGGCAAGTTCACTACTTCTCGAATCTTACTACAGCTATGAGGTCCTCTCGGCGC 179

OY 184 cgttaacagctgaccttattgtgaagaanaaacaagtcatttttctgtaaatgcatctg 243
    |||
Db 180 CAGTGACAGTGAATTCATCGAAGAAAGAAAGAGTCCACTTTTGTGATGAGGTCGTG 239

OY 244 atgtcacaagaatgtgaagaaccatttaaatgtgagcgggaattccatgcatgtctgtg 303
    |||
Db 240 ATGTGACAGTGTCAAAACGCAATTTAAATGTGAGCAGAAATTCATTAACTGTCGATGA 299

OY 304 cagacgtgaagatcttactacaagcagatttccaaacagacagtcagcccccagagct 363
    |||
Db 300 ACAAGGTGAGAGACTTAATTGAACGACGACTTTCAATGACACGGTCAGCCCCGCGCT 359

OY 364 ccgcatcgtactatgaacagiatcactcactaaatgaactatcttctgtgataagatta 423
    |||
Db 360 CCGCTTCATTACTATGAGCAGATCATCGCTAAATGAATCTATTCCGTGATGAGATCA 419

OY 424 taactgagagcatctctgtatatagtcttaacaaaatccaatgtgactctcaattgagaag 483
    |||
Db 420 TAACTGAACACACATCTGACATGCTCCAGAAAATCTACATGAGATCAATTCAGAGAGT 479

OY 484 accacatctatgtttaaaggcttctggaagaagaacaacagccaaaatgcatatgga 543
    |||
Db 480 ACCCATTTATGTTTAAAGTCTCAGAGAAAGAAAGAAATCAAAATCCATCTGGA 539

OY 544 ttgactgtgaatccatgacagagaatgagatctctctgtcttctgtctgtgtgtgtaag 603
    |||
Db 540 TCGACTGTGAATCCATGCGCAGAGAAATGATTTGACCGTCTTGTGTTGGTTCAATAG 599

OY 604 gc----- 605
    ||
Db 600 GCTTACGTGACAAATTCATGCGGAAAGAAATCTGTATACCAGACTTCTGAGCGACGTGG 659

OY 606 -----catac 612
    ||
Db 660 ATTCTACATCATGCGCGGTATGAAACGTGATGAGTGGCTATGACTACGTGGAAAAGAAATC 719

OY 613 gaatgtggaagaagaacagcttcttctatgcgaacatcatgcatggaacagacctga 672
    |||
Db 720 GAATGTGAGAGAAACCGCTCTCTCACAAGAAACACCGCTGCGTGGCACAACCTGA 779

OY 673 atagcaactgttctccaacacactggtgtgaggaagtgatccagttctcctcgcg 732
    |||
Db 780 ACAGGAACCTTCCTCCAAACACTGTGTGTGAAGAAAGGTGCGTCAAGTTCCTCTGCTCG 839

OY 733 aaacctactgtgaacttactctgtagtcagaaacagagaagtgaagtcagttgttct 792
    |||
Db 840 AATCTACTGTGAGACTTATCTGAGTCTGAGCCAGAGGGAAGGAGAGTGGTACTCTT 899

OY 793 tgaagaagaataataacacagattaaagacatacatgcatgcatctactccagcata 852
    |||
Db 900 TGAGAAAGAAATATGACCACTATTAACCTTACACACTATGACCTCATCTCCCAACAAA 959

OY 853 tagtttccatctctctatacacgaagttaaagcaagaacataggaacttctctag 912
    |||
Db 960 TACTGTTTCCCTATTCCTATTAACAGAAAGCAAAAGCAAGCAAGAAAGTGTCTTAG 1019

OY 913 tagcagatgaagaagtcgtgtctatgtgacaacaaactagtaaaaaatccaggtatcacatg 972
    |||
Db 1020 TGGCCACGAGAAAGCTGCGGCAATTAAGATTAATAAACAACAGGTATACACACAGC 1079

OY 973 gccatgtcagaacacttactagctcctgtgaggtgtgagcagatgtgactatgatt 1032
    |||
Db 1080 GCAATGCTCAGAAAGTTTATATCTACTCTCTGAGGTTCTGACAGATGATCTATGATT 1139

OY 1033 tgggcatcaaatatttgt----- 1051
    |||
Db 1140 TGGGCAATCAATATTCGTTTACATTGAGCTCCGAGATACAGGACAGATACGATTCCTGGC 1199

```


QY	1052	-----taactcaaacccacccctctagagaagcctttgcgcgtctctcttaaatag	1100.
Db	1200	TGCCCTGAGAGAAATCAACCCACTTTGTGCAAGAAAGCTTTGGCCGCCATCTCTAAATAAG	1259
QY	1101	cttgcgcgtcaatcattaggaattgttaatgcctccctgtatttcaatcgtctccgtatttt	1160
Db	1260	TTTGGCAGTCTCAATCAGGAACACTTAATGCC-----TAACCTCCGCTCTCATTAATTTTT	1313
QY	1161	aatttactgattccgaagaacaaatcatatgtaacagat--tattttaagttttatccc	1218
Db	1314	ATTTATATGATTTTTCAGCAACACTTAACGTTGTCATTAAGTCTTAAGTGAATCAAGTTTCC	1373
QY	1219	gtagtttgcataaaagat	1236
Db	1374	TTGTGTTTGTGTTGAGAAAT	1391
RESULT	8		
LOCUS	AF164524	1490 bp	linear
DEFINITION	Mus musculus thrombin-activatable fibrinolysis inhibitor (Tafi)		
ACCESSION	AF164524		
VERSION	AF164524.1	GI:7416966	
KEYWORDS			
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1490) Marx,P.F., Wegenaar,G.T.M., Reijerkerk,A., Tiekstra,M.J., van Rossum,A.G.S.H., Gebbink,M.F.G.B. and Meijers,J.C.M.		
TITLE	Characterization of mouse thrombin-activatable fibrinolysis inhibitor		
JOURNAL	Thromb. Haemost. 83 (2), 297-303 (2000)		
MEDLINE	20201996		
REFERENCE	2 (bases 1 to 1490) Marx,P.F., Wegenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.		
AUTHORS	Direct Submission		
TITLE	Submitted (01-JUN-1999) Haematology, UMCU, Heidelberglaan 100, Utrecht 3584 CX, The Netherlands		
JOURNAL			
FEATURES	Location/Qualifiers		
source	1..1490		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:477035"		
	1..1490		
	/gene="Tafi"		
	84..1352		
	/gene="Tafi"		
	/function="carboxypeptidase"		
	/codon_start=1		
	/product="thrombin-activatable fibrinolysis inhibitor"		
	/protein_id="AA662385.1"		
	/db_xref="GI:7416967"		
	/translation="MKLHGGLILVAIIIEQGFAPFSGVIALAPRTSROVLQNLNITTEYVLMOPVTAIEFIEKKVEHFVNASVDYSVKALINVSRIPTNLMNVNDELIEOOFNDYSPRASASVYEOYSLEIYSIVITIEHDMOKIYISSEPKYELIYVYKSGEORIKNAIINIDCGTARRENIISPAFCMTFTGYTOFGKENTLTRILRHADFTYMPVNVVDSEYDTWKKNNRWNRNRSAAHNNRCVGDILNNEFASKHMCGEKGAASSSCSEYVCGYIYVDEPKAVADFLRNIDHINIKAYISMHSOOILFPYSYNRKSKDHDELISVASSEVRAIESINIKNTRYTHGSGESLTLAAGSGDDWIYDGIKYSFTIELDTGRVGGVFLPFRYIKPFCAEELAIASIKIYWHVINT"		
BASE COUNT	438 a	337 c	322 g
ORIGIN	337 c	322 g	393 t
Query Match	41.7%;	Score 656;	DB 10;
Best Local Similarity	72.2%;	Pred. No. 1.9e-141;	
Matches 1009;	Conservative	0;	Mismatches 215; Indels 174; Gaps 5;
4	aaatctgctgttggatgaagccttgcagccttcgagctctgtataccatgttctctct	63	

D	b	70	AAGTAC	CTGTTGGGATGACGCTTATGCGCCTTGGAATCTCGTAAGCCATCATCTCTTA--	127
O	y	64	gtgaagca	gcatgctctgcgcgttcccaagtgccaagttcagctgctctcccaagact	123
D	b	128	-TGAGC	ACATGSGCTTGCGCTTTAGAGTGGCCAGGTTTATCTGCTTCCAAAGAACT	186
O	y	124	ctaggcaag	tccaagttcctacagaatcttaactacacataatgaatgtgtctctgcagc	183
D	b	187	CCAGGCA	GTTCAACTACTTCAGAACTTCTACTCAACAGTATGAGGTCGTCTCTGGCAGC	246
O	y	184	cggtaa	cgcgcgacctattgttgaagaaaaacaagccatltttttgtaaatgacctg	243
D	b	247	CAGTGAC	GCCTGAATTCATTCGAGAAAGAAAAGAAAGTCCACTTTTGTGTAAGCCGTCTG	306
O	y	244	atgtgaca	atgtaaagcccatctaagtgtgaagcggatcccatcgaatgtctgtcg	303
D	b	307	ATGTGCA	ACAGTGTCAAAAGCGATTAAATGTGACAGAAATTCATTAAAGTTCTATTA	366
O	y	304	cagaatg	tggaagatcttattcaacagcagaattccaacgcacagtcagcccccgacct	363
D	b	367	ACAAGTG	AGAGACTTATTGAACAGAGACTTTCATGTACAGCGTCAGGCCCCGGCGCT	426
O	y	364	ccgatcgt	tactatgaacagtatatactcaactaatgaatcatctattcttgatagaatt	423
D	b	427	CCGCTT	CTACTATGAGAGATACTCGCTAAATGAAATCTATTCCTGGATTTGAAGTCA	486
O	y	424	taactga	aagcaccctcgaatgtgttcaaaaaaccacatlgatctctcaatttgagaagt	483
D	b	487	TAACTGA	ACAGCATTCCTGACATGCTCCAGAAATCTTCATCGGATCTCATTTGGAAGAT	546
O	y	484	accaccta	tgctttaaaggttctgtgaaagaacacaaacgcacaaataatgacata	543
D	b	547	ACCCACT	TATGTTTAAAGGTCACAGAAAGAAAGAAATCAAAATCAAAATGCGATCGGA	606
O	y	544	ttagctgt	ggaatccabgcagagaaatgagatctcctgcctctctgtctgtgttgatag	603
D	b	607	TGCACTG	GGAATCCATGCCAGAAATGAGATTCACCTGCTTCTGTTGTGGTTATAG	666
O	y	604	gc	-----	605
D	b	667	GCTACG	TACACATTCCTCATGGAAGAAATCTGTATACCAACATCTTGAGCACGTGG	726
O	y	606	-----	-----cataatc	612
D	b	727	ATTTCAT	CATCATGCGCGGTATGAACGATGCGTATGACTACAGCTGGAAGAAAGATC	786
O	y	613	gaatgtg	aaagaagaccggtctcttctatctgcgaacaatcatitgcatcgtgaacagctga	672
D	b	787	GAAATG	AGAGAAAGCGCTCTCTCAACAAGAACCCCTCGTGGCCACAGACTCGA	846
O	y	673	atagaact	gtgtcccaaacacgtggtgtaggaaggtgcatcaagtctctcagtcg	732
D	b	847	ACAGGA	ACTTGGCTTCACAAACACTGgtgtgGAAAGGTGTCAAGTCTCTCGTCTG	906
O	y	733	aaacctact	gtggaacttatcctcgtagtcagaacacagaagtgaagcgagttgctagtctt	792
D	b	907	AAACTAC	TGTGGACTTATCTCTGAGTGTGAGCCAGAGGTGAAGGCGAGTGGCTACTGT	966
O	y	793	tgaaga	aaataccaacagattaaagcatatacagcatgcatctactactcccaagata	852
D	b	967	TGAGAGA	AAATAATCGACCCACTTTAAAGCTTACATCACTATGCACTCATCTCCACAA	1022
O	y	853	tagtgtt	cacatctccatacaagaagttaaagcaaaagccatlgagaaactgtctctag	912
D	b	1027	TACTGT	TTCCTATTCTATACAGAAAGCAAAACAGACACAGAAACTGTCTCTAG	1088
O	y	913	taggcaag	taagcagttctgtctatctgaacaaatagtaaaaaataccagtatatacatg	972
D	b	1087	TGGCCAG	GGAAGCAAGTGTGTCGAFTGAAGTATTAATAAAACACAGGTACACACAG	114
O	y	973	gccatgct	caagaaactataactagctctgtgaagtgaggacagatgtgatgtatgatt	103
D	b	1147	GCAGTGC	CTCGAAAGTTTATATCTAGTCTCTGAGGCTTTCAGATGATGATCATAGATT	1201

[illegible]

misc_feature	/clone_11b="RPT-11.1"	complement(1..95)	/note="match: GSS: Em:AQ266238"
repeat_region	624..851		
repeat_region	/note="LIPAB repeat: matches 5955..6162 of consensus"		
repeat_region	1331..1611		
misc_feature	/note="LIMC/D repeat: matches 5541..5822 of consensus"	complement(1912..2370)	
repeat_region	2288..3195	/note="match: GSS: Em:AQ335999"	
repeat_region	/note="LIME3 repeat: matches 5203..6161 of consensus"	3676..4108	
repeat_region	/note="Char11a1b repeat: matches 2..463 of consensus"	4453..4871	
misc_feature	/note="I2 repeat: matches 1151..1600 of consensus"	complement(4551..5058)	
repeat_region	/note="match: GSS: Em:AQ662290"	5674..5855	
repeat_region	/note="ME558 repeat: matches 1..189 of consensus"	6139..6317	
repeat_region	/note="AluIO repeat: matches 16..193 of consensus"	6318..6416	
repeat_region	/note="AluIO repeat: matches 194..283 of consensus"	9114..9332	
repeat_region	/note="ME546C repeat: matches 23..333 of consensus"	10838..11072	
repeat_region	/note="HA11 repeat: matches 571..819 of consensus"	11095..11216	
repeat_region	/note="LIMEC repeat: matches 1471..1593 of consensus"	11771..12046	
repeat_region	/note="AluIO repeat: matches 12..284 of consensus"	12651..12732	
repeat_region	/note="41 copies 2 mer aa 63% conserved"	12920..13482	
repeat_region	/note="LIM4 repeat: matches 2645..3199 of consensus"	13549..13850	
repeat_region	/note="AluIO repeat: matches 3..305 of consensus"	13884..14086	
repeat_region	/note="LIM4 repeat: matches 3359..3566 of consensus"	14322..15872	
repeat_region	/note="LIMEC repeat: matches 2383..3622 of consensus"	15873..16175	
repeat_region	/note="AluY repeat: matches 1..303 of consensus"	16176..16335	
repeat_region	/note="LIMEC repeat: matches 3622..3777 of consensus"	16316..16419	
repeat_region	/note="LIM49 repeat: matches 6158..6264 of consensus"	16425..18501	
repeat_region	/note="LIMEC/D repeat: matches 3729..5776 of consensus"	18517..18700	
repeat_region	/note="LIMC1 repeat: matches 6134..6318 of consensus"	18790..18835	
repeat_region	/note="23 copies 2 mer aa 78% conserved"	19057..19377	
repeat_region	/note="LIME repeat: matches 5607..5930 of consensus"	20066..20261	
repeat_region	/note="LIMEC/D repeat: matches 5280..5504 of consensus"	20263..21196	
repeat_region	/note="LIME1 repeat: matches 5133..6163 of consensus"	21376..21509	
repeat_region	/note="LIMD3 repeat: matches 6602..6740 of consensus"	21570..21822	
misc_feature	/note="MLT1b repeat: matches 227..490 of consensus"	21578..22061	
repeat_region	/note="match: GSS: Em:AQ697967"	22450..22616	
repeat_region	/note="AluIO repeat: matches 137..303 of consensus"	23172..23466	
repeat_region	/note="AluSX repeat: matches 1..305 of consensus"	23539..23673	
repeat_region	/note="LIMB4 repeat: matches 6043..6182 of consensus"	24142..24404	
repeat_region	/note="MER33 repeat: matches 53..324 of consensus"		

```

repeat_region 24427..24464
/note="MER5B repeat: matches 1..38 of consensus"
24493..24551
/note="L2 repeat: matches 2606..2659 of consensus"
repeat_region 25199..25476
/note="AluSg repeat: matches 11..287 of consensus"
26044..26421
/note="MER63B repeat: matches 1..435 of consensus"
26438..26592
/note="MER5A repeat: matches 2..149 of consensus"
26821..28021
/note="L1MB8 repeat: matches 4920..6157 of consensus"
28733..28774
/note="21 copies 2 mer aa 76% conserved"
misc_feature complement(30583..31098)
/note="match: GSS: Em:AQ0507270"
31066..31450
/note="match: GSS: Em:AQ094491"
32825..33131
/note="AluY repeat: matches 1..297 of consensus"
34745..35053
/note="AluY repeat: matches 2..306 of consensus"
35297..35616
/note="AluSc repeat: matches 1..297 of consensus"
complement(37235..37677)
/note="match: GSS: Em:AQ058283"
38768..39332
/note="Cpg island"
evidence=not_experimental
complement(39645..40061)
/note="match: STS: Em:G05971"
39646..39826
/note="match: STS: Em:G59992"
39759..39820
/note="2 copies 31 mer 100% conserved"
41211..41498
/note="AluSx repeat: matches 1..288 of consensus"
41830..41895
/note="MER91B repeat: matches 1..66 of consensus"
45630..45734
/note="MIR repeat: matches 148..256 of consensus"
45811..46027
/note="MIR repeat: matches 39..261 of consensus"
46028..46069
/note="L4 copies 3 mer cat 100% conserved"
46245..46473
/note="match: STS: Em:HSPH32C9"
46546..46628
/note="MER33 repeat: matches 238..324 of consensus"
46629..46907
/note="AluSx repeat: matches 1..280 of consensus"
46908..47163
/note="MER33 repeat: matches 4..228 of consensus"
47214..47332
/note="MIR repeat: matches 12..137 of consensus"
47409..47662
/note="L1MB4 repeat: matches 5924..6181 of consensus"
47817..48194
/note="L1MB4 repeat: matches 1..390 of consensus"
48463..48492
/note="15 copies 2 mer tt 90% conserved"
48511..48946
/note="match: STS: Em:G36957"
48722..48834
/note="MIR repeat: matches 35..138 of consensus"
49106..49234
/note="L2 repeat: matches 2605..2750 of consensus"
49628..50196
/note="L1PA16 repeat: matches 5577..6144 of consensus"
50428..50717
/note="AluSx repeat: matches 1..292 of consensus"
complement(50645..50888)
/note="match: GSS: Em:AQ105432"
misc_feature

```

```

misc_feature 51632..52077

Query Match 30.2%; Score 475.4; DB 9; Length 58097;
Best Local Similarity 97.6%; Pred. No. 9..9e-100;
Matches 493; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 1051 ttacatcaaacccactgtagagaagctttgcgcgtctctctaaatagctggcatgt 1110
|||||
DB 40137 ttacatcaaacccactgtagagaagctttgcgcgtctctctaaatagctggcatgt 40078

QY 1111 ctttagaatgtttaatgccccgatttatcatcttgccttgatatttaattactga 1170
|||||
DB 40077 ctttagaatgtttaatgccccgatttatcatcttgccttgatatttaattactga 40018

QY 1171 ttccagcaagaccacaatcatgtatcatatttttaagtttaccgttagtttgata 1230
|||||
DB 40017 ttccagcaagaccacaatcatgtatcatatttttaagtttaccgttagtttgata 39958

QY 1231 aaagatttccatttcccttggtctgtcagagaacctaaatagtgtaacttgcaat 1290
|||||
DB 39957 aaagatttccatttcccttggtctgtcagagaacctaaatagtgtaacttgcaat 39898

QY 1291 ggcaagactagggtctatgtcttttaacccttaaaaaaatgttaagttcattacc 1350
|||||
DB 39897 ggcaagactagggtctatgtcttttaacccttt-AAAAAAAAATGTAAAGCTTAGTTACC 39839

QY 1351 tacttttcttcttatttcgaagcttgaactgacatcgaactttcgaagcttgac 1410
|||||
DB 39838 tacttttcttcttatttcgaagcttgaactgacatcgaactttcgaagcttgac 39779

QY 1411 tagcatctcaagaagtttaatcaagaatcatcgaagcttgatctgactactca 1470
|||||
DB 39778 tagcatctcaagaagtttaatcaagaatcatcgaagcttgatctgactactca 39719

QY 1471 acaaaaggaaagggtgtcgaagaatcatatgaatttcgtccaaatttcaataatt 1530
|||||
DB 39718 acaaaaggaaagggtgtcgaagaatcatatgaatttcgtccaaatttcaataatt 39659

QY 1531 tcttcttctctcttaaaaaa 1555
|||||
DB 39658 tcttcttctctctttagaaattacaa 39634

RESULT 12
LOCUS G05971 437 bp DNA linear STS 19-OCT-1995
DEFINITION human STS WI-7773, sequence tagged site.
ACCESSION G05971
VERSION G05971.1 GI:859216
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STS derived from sequences in dbEST and the Unigene collection.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 437)
REFERENCE
AUTHORS Hudson, T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
Mapped ESTs
JOURNAL Unpublished
COMMENT
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: TTAATTACTGATTCACGACGACC
Primer B: TCTGACACCTTCCTTTTG
STS size: 333
PCR Profile:

```

Presoak:
 Denaturation: 56 degrees C
 Annealing: 56 degrees C
 Polymerization: 35
 PCR Cycles: 35
Thermal Cycler:
 Protocol:
 Template: 10 ng
 Primer: each 5 pm
 dNTPs: each 4 mM
 Tag Polymerase: 0.025 units/ul
 Total Vol: 20 ul
Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3

FEATURES
 Location/Qualifiers
 source 1..437

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="740_H-5; 760_B-8; 762_A-1; 773_H-2; 798_H-1;
 802_G-5; 872_F-3; 922_E-2; 937_C-7; 955_G-7; 817_F-11;
 846_A-11; (741-748)_A-5"
 33..365
 STS primer_bind complement(346..365)
 BASE COUNT 109 a 84 c 60 g 156 t 28 others
 ORIGIN

Query Match 26.0%; Score 409; DB 11; Length 437;
 Best Local Similarity 97.8%; Pred. No. 2.8e-84;
 Matches 409; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Oy 1127 tggccgatttcatcattcgtctcgtatcttaattactgattccagaagaccaaa 1186
 Db 1 tggcccgatttcatcattcgtctcgtatcttaattactgattccagaagaccaaa 60
 Oy 1187 tcaattgatacagatttttaagtttaccgtagtttgataaagatttctctatc 1246
 Db 61 tcaattgatacagatttttaagtttaccgtagtttgataaagatttctctatc 120
 Oy 1247 cttggttcgtcagaagaccataaagtcctacttgcacatgaagcagactaaggttca 1306
 Db 121 cttggttcgtcagaagaccataaagtcctacttgcacatgaagcagactaaggttca 180
 Oy 1307 tgccttttacccttcaaaaaaattgttaaaagtcctacttgcacatgaagcagactaaggttca 1366
 Db 181 tgccttttacccttcaaaaaaattgttaaaagtcctacttgcacatgaagcagactaaggttca 240
 Oy 1367 ttcgagcgtttgactgacatcctcaagcaacttgcagcgtttgactgacatcctcaagca 1426
 Db 241 ttcgagcgtttgactgacatcctcaagcaacttgcagcgtttgactgacatcctcaagca 300
 Oy 1427 gtttaacaaagatcatctcagctgatactgatactcactcaaaaaggagggttg 1486
 Db 301 gtttaacaaagatcatctcagctgatactgatactcactcaaaaaggagggttg 360
 Oy 1487 tcagaagatcaataaagattctgtctccaaatttcaaaaatttcttcttctt 1544
 Db 361 tcagaagatcaataaagattctgtctccaaatttcaaaaatttcttcttcttctt 418

RESULT 13
 G59992/c G59992 181 bp DNA linear STS 30-MAR-2000
 LOCUS SHG-130613 Human Homo sapiens STS genomic, sequence tagged site.
 DEFINITION G59992
 ACCESSION G59992
 VERSION G59992.1 GI:6125161

KEYWORDS
 STS.
SOURCE
 human.
ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
 1 (bases 1 to 181)
 Olivier, M. and Cox, D.R.
 Unpublished, Olivier, M., Cox, D.R. (2000)
 Unpublished

AUTHORS
 Olivier, M. and Cox, D.R.
TITLE
 Unpublished, Olivier, M., Cox, D.R. (2000)
JOURNAL
 Unpublished
COMMENT
 Contact: Michael Olivier, David R. Cox
 Stanford Human Genome Center
 Stanford University School of Medicine
 4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
 Tel: (650) 320-5800
 Fax: (650) 320-5801
 Email: oliviereshg@stanford.edu
 Primer A: TGTACTTCTGACCACTTCTCTT
 Primer B: TTTCAGCTTTGACTAGCCATCT
 STS size: 101
 PCR Profile:
 Initial incubation: 95 degrees C for 10 minutes
 Denaturation: 94 degrees C for 30 seconds
 Annealing: 60 degrees C for 30 seconds
 Polymerization: 72 degrees C for 23 seconds
 PCR Cycles: 30
 Thermal Cycler: Perkin Elmer 9700

Protocol:
 Template: 25 ng
 Primer: each 1 uM
 dNTPs: each 200 uM
 Amplitag Gold Polymerase: 0.07 units/ul
 Total Vol: 5 ul

Buffer:
 MgCl2: 2.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 8.3

Finished human sequence in NCBI. STSs designed and developed at the
 Stanford Human Genome Center.

FEATURES
 Location/Qualifiers
 source 1..181

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="13"
 /clone_id="Human"
 STS primer_bind complement(125..147)
 primer_bind 47..147
 primer_bind 47..69
 BASE COUNT 56 a 31 c 40 g 53 t 1 others
 ORIGIN

Query Match 11.0%; Score 173.6; DB 11; Length 181;
 Best Local Similarity 97.2%; Pred. No. 8.4e-30;
 Matches 176; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1363 gatttgcagctttgactgacatcctcaagaacttgcagcgtttgactgacatcctca 1422
 Db 181 gatttgcagctttgactgacatcctcaagaacttgcagcgtttgactgacatcctca 122
 Oy 1423 gaaagtttaacaaagatcatctcagctgatactgatactcactcaaaaaggagg 1482
 Db 121 gaaagtttaacaaagatcatctcagctgatactgatactcactcaaaaaggagg 62
 Oy 1483 gtcgtcagaagatcaataaagattctgtctccaaatttcaaaaatttcttcttct 1542
 Db 61 gtcgtcagaagatcaataaagattctgtctccaaatttcaaaaatttcttcttcttcc 2
 Oy 1543 t 1543

```

Db      1 T 1

RESULT  14
AL137141
LOCUS   AL137141
DEFINITION Human DNA sequence from clone RP11-139H14 on chromosome 13 contains the LCP1 (lymphocyte cytosolic protein 1 (L-plastin)) gene, part of the gene for CPB2 (carboxypeptidase B2 (plasma)), ESTs, STSs, GSSs and a CpG island, complete sequence.
ACCESSION AL137141
VERSION   AL137141.10
KEYWORDS  HTG; carboxypeptidase B2; CPB2; CpG Island; cytosolic; L-plastin; LCP1; lymphocyte.
SOURCE    human.
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 180680)
AUTHORS   Tromans,A.
JOURNAL   Direct Submission
           Submitted (24-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
           requests: clonerequests@sanger.ac.uk
           On Jun 3, 2000 this sequence version replaced gi:7523347.
           During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
           This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
           The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; SW:, SWISSPROT; Tr:, TREMBL; Wp:, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep
           This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr13
           RP11-139H14 is from the library RPCI-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/VECTOR: PBAC3.6
           This sequence is the entire insert of clone RP11-139H14 The true left end of clone RP11-106H11 is at 103183 in this sequence.
FEATURES
SOURCE   1..180680
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /chromosome="13"
          /clone="RP11-139H14"
          /clone_lib="RPCI-11.1"
          453..898
          /note="match: GSS: Em: A0601126"
          465..544
          /note="40 copies 2 mer ta 75% conserved"
          548..754
          /note="LIMC4 repeat: matches 6571..6787 of consensus"
          1297..1361
          /note="MIR repeat: matches 82..146 of consensus"
          2531..2642
          /note="LIM4 repeat: matches 4637..4744 of consensus"
          2674..3981
          /note="LIM4 repeat: matches 2110..3419 of consensus"
          5917..6306
          /note="MSTB repeat: matches 1..426 of consensus"

repeat_region 6665..7452
               /note="LIMC repeat: matches 2400..2858 of consensus"
repeat_region 7453..7742
               /note="AluYo repeat: matches 1..290 of consensus"
repeat_region 7743..7940
               /note="LIMC repeat: matches 2209..2400 of consensus"
repeat_region 7971..8339
               /note="LIMC repeat: matches 1855..2223 of consensus"
misc_feature 8282..8613
               /note="match: GSS: Em: A082919"
repeat_region 8454..8548
               /note="AluS repeat: matches 1..95 of consensus"
repeat_region 8538..8761
               /note="L1 repeat: matches 2193..2404 of consensus"
repeat_region 8754..9754
               /note="L1M1 repeat: matches 1156..2544 of consensus"
repeat_region 9764..11803
               /note="L1PA13 repeat: matches 4126..6144 of consensus"
repeat_region 11804..12109
               /note="AluY repeat: matches 1..307 of consensus"
repeat_region 12110..13463
               /note="L1PA13 repeat: matches 2741..4126 of consensus"
repeat_region 13464..13816
               /note="THE1B repeat: matches 1..364 of consensus"
repeat_region 13817..14231
               /note="L1PA13 repeat: matches 2332..2741 of consensus"
repeat_region 14244..14735
               /note="L1 repeat: matches 3036..3533 of consensus"
repeat_region 14739..15072
               /note="LIMC1 repeat: matches 5928..6260 of consensus"
repeat_region 15087..16456
               /note="LIMC repeat: matches 1271..2076 of consensus"
repeat_region 16457..16559
               /note="L1M4 repeat: matches 1271..2076 of consensus"
repeat_region 16560..17236
               /note="L1M3 repeat: matches 1..103 of consensus"
repeat_region 17243..17307
               /note="LIMC repeat: matches 588..1271 of consensus"
repeat_region 17318..17480
               /note="LIMD3 repeat: matches 7682..7739 of consensus"
repeat_region 19367..19524
               /note="MIR repeat: matches 20..187 of consensus"
repeat_region 19561..19592
               /note="LIMB repeat: matches 1032..1190 of consensus"
repeat_region 19945..20115
               /note="16 copies 2 mer ct 84% conserved"
repeat_region 20116..20414
               /note="MSTB repeat: matches 1..169 of consensus"
repeat_region 20425..20725
               /note="AluY repeat: matches 1..299 of consensus"
repeat_region 20726..20964
               /note="AluS repeat: matches 1..305 of consensus"
repeat_region 20966..21618
               /note="MSTB repeat: matches 177..426 of consensus"
repeat_region 21629..21931
               /note="MSTB-internal repeat: matches 2..667 of consensus"
repeat_region 21932..22073
               /note="MSTB repeat: matches 233..553 of consensus"
repeat_region 22075..22480
               /note="MSTB-internal repeat: matches 1501..1650 of consensus"
repeat_region 22481..22589
               /note="MSTB repeat: matches 1..426 of consensus"
repeat_region 24229..24707
               /note="FLAM-A repeat: matches 1..113 of consensus"
misc_feature 25947..26255
               /note="match: STS: Em: T03078"
repeat_region 26647..26749
               /note="AluYo repeat: matches 1..311 of consensus"
repeat_region 27045..28334
               /note="L2 repeat: matches 1161..1267 of consensus"
repeat_region 28335..28656
               /note="LIM2 repeat: matches 413..1676 of consensus"
repeat_region 28335..28656
               /note="AluSg repeat: matches 1..310 of consensus"

```

```

repeat_region 28657..28888
/note="LIM2 repeat: matches 1676..1910 of consensus"
repeat_region 28882..28967
/note="LIM1 repeat: matches 4801..4906 of consensus"
repeat_region 29015..29706
/note="LIP3 repeat: matches 5455..6146 of consensus"
repeat_region 29707..29918
/note="LIM1 repeat: matches 4892..5103 of consensus"
repeat_region 29924..30246
/note="LIM1 repeat: matches 5212..5403 of consensus"
repeat_region 31065..31296
/note="L2 repeat: matches 2477..2706 of consensus"
repeat_region 31301..37567
/note="LIP3 repeat: matches 144..6143 of consensus"
repeat_region 37667..37730
/note="MER20 repeat: matches 159..218 of consensus"
repeat_region 38334..38909
/note="MER82 repeat: matches 6..605 of consensus"
misc_feature 41615..42079
/note="match: GSS: EM: N0583743"
complement(42645..42669)
/note="match: STS: EM: G24043"
repeat_region 43129..43189
/note="L2 repeat: matches 2648..2708 of consensus"
repeat_region 43661..43874
/note="MIR repeat: matches 23..253 of consensus"
repeat_region 46480..46765
/note="LIP3 repeat: matches 1..287 of consensus"
repeat_region 46766..46952
/note="L1 repeat: matches 5164..5370 of consensus"
repeat_region 47102..47417
/note="LIP3 repeat: matches 1..311 of consensus"
repeat_region 47518..47759
/note="MIR repeat: matches 13..262 of consensus"
repeat_region 47887..47928
/note="L2 repeat: matches 545..679 of consensus"
repeat_region 49439..49482
/note="MIR repeat: matches 107..150 of consensus"
repeat_region 49517..49757
/note="MIR repeat: matches 6..262 of consensus"
repeat_region 50473..50594
/note="L2 repeat: matches 545..679 of consensus"
repeat_region 50599..50898
/note="LIP3 repeat: matches 1..304 of consensus"
repeat_region 50906..51438
/note="LIM3 repeat: matches 7223..7739 of consensus"
repeat_region 51480..52374
/note="L2 repeat: matches 693..1543 of consensus"
repeat_region 52375..52536
/note="LIP3 repeat: matches 5997..6155 of consensus"
repeat_region 52537..52577
/note="L2 repeat: matches 1543..1581 of consensus"
repeat_region 52578..52882
/note="LIP3 repeat: matches 1..305 of consensus"
repeat_region 52883..53971
/note="L2 repeat: matches 1581..2750 of consensus"
repeat_region 56191..56608
/note="MIR repeat: matches 8..426 of consensus"
repeat_region 57138..57237

```

Query Match 8.1%; Score 128; DB 9; Length 180680;
Best Local Similarity 87.5%; Pred. No. 2.2e-19;
Matches 140; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Oy 133 ttcaagttctacagatcttactacacataatgagatgtctctcgtgagccggttaacag 192
|||||
Db 167852 TTTTAAATTTTTCATTTTGGCCCTTAACCTGATGATGTTCTCTGCGACGCGTAACAG 167911

Oy 193 ctgacctatctgtgaagaaacagatccatctttttgttaaatgcatctgcatgctgaca 252
|||||
Db 167912 CTGACCTTATGCTGAAGAAAAAAGATCCATTTTCTGTAATGATCATCTGATGACACA 167971

Oy 253 atgtgaagcccatcttaaatgtgagcggaattcatgacag 292

```

|||||
Db 167972 ATGTGAAGCCCATTTAATGTAGCGGAAATTCATGACG 168011

RESULT 15
LOCUS AR072911 927 bp DNA linear PAT 28-AUG-2000
DEFINITION Sequence 5 from patent US 5948668.
ACCESSION AR072911
VERSION AR072911.1 GI:9999674
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 927)
AUTHORS Hartman,J., Fulda,N., Mendelovitch,S. and Gorecki,M.
TITLE Production of enzymatically active recombinant carboxypeptidase B
JOURNAL Patent: US 5948668-A 5 07-SEP-1999;
FEATURES
source location/Qualifiers
1..927
/organism="unknown"
BASE COUNT 253 a 226 c 213 g 235 t
ORIGIN

```

Query Match 7.9%; Score 124; DB 6; Length 927;
Best Local Similarity 57.9%; Pred. No. 2.3e-18;
Matches 278; Conservative 0; Mismatches 195; Indels 7; Gaps 3;

Oy 613 gaatctgagaagaagacgcttcttctatgacacatcatctatcaggaacagctga 672
|||||
Db 362 GAATGTGAGAAAGAACCCGCTCTACTATGCTGGAGTTCCTGTTGGGTGATGACCCCA 421

Oy 673 atagcaattgtctccaacactggtgtaggaaggtgcatccatctcctgctcg 732
|||||
Db 422 ACAGGATTTTAATGCG--TGCTGTGTGAGAGTGGAGGCTTCTCGAGTCCCTGCTTG 478

Oy 733 aaactactgtgacttaccctcagagtcagagacagaagtggaagcgagtgacttct 792
|||||
Db 479 AAATTTACTGTGACACGCCCCAGAGCTCTGAAAAGAGAAAGGCCCTTGCGAGATTTC 538

Oy 793 tgaagaagaatataacacagataaagacalacatcagcatgcatctactaccagcata 852
|||||
Db 539 TCGCAACAAACCTCTCCACCATCAAGGCTACCTGACCATCTCATCTACACATGA 598

Oy 853 tagtgcttcattctctatatacagaagtaagaagaagccatgaggaactgtctag 912
|||||
Db 599 TGCTCTACCTTACTCTCTATGACTACAACTGCTGGAATCTGAGAAATGGAATGCC 658

Oy 913 tagcagtgagagcagtcgctgctatgacaaaactaglaaaaataccaggtatcacatg 972
|||||
Db 659 TGTGCAAGGTGCGCGCAAGAGAGCTTCC--ACTCTGCAATGGCACCAAGTACACATATG 715

Oy 973 gccatgctcagaacacttatacctagctcctctgaggttgaggagcatgatatgatt 1032
|||||
Db 716 GCCCAGAGGCTACCAACATCTATCTCTGCTGGGGATCTGACGAGTGGTCTTATATC 775

Oy 1033 tgggcatacaatattgctttacatcaaacacacacatgtagagaagccttgccgctctc 1092
|||||
Db 776 AGGGAATCAATATTCCTTACCTTTGA-ACTCCGGATACAGGCTTCTTGGGCTTCTC 834

Search completed: September 18, 2002, 16:05:39
Job time: 2842 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2002, 15:24:02 ; Search time 210.94 Seconds

(without alignments)
12803.197 Million cell updates/sec

Title: US-09-980-881-1

Perfect score: 1573
Sequence: 1 agaaaaatcgttgggatg.....aaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802:*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1573	100.0	1573	21	AAC81962 Human brain carbox
2	1372	87.2	1749	17	AA11671 Human plasma carbo
3	1372	87.2	1749	18	AA162846 Human plasma carbo
4	1370.4	87.1	1749	14	AA041001 Human plasma carbo
5	918	58.4	1272	20	AAV74302 Human plasma carbo
6	853.8	54.3	1400	21	AA18005 Lung cancer associ
7	324	20.6	333	20	AAH85957 Human single nucle
8	324	20.6	333	20	AAH87390 Human single nucle
9	241.6	15.4	251	19	AAH11374 Human biallelic po

10	124	7.9	927	17	AAT35760 Rat mature carboxy
11	123.4	7.8	921	16	AA090601 Porcine carboxypep
12	123.4	7.8	1215	16	AA090600 Porcine Tyr-His-Me
13	120	7.6	121	20	AAH85710 Human single nucle
14	120	7.6	121	20	AAH85711 Human single nucle
15	109	6.9	1263	19	AAV41795 Human pancreatic c
16	107.4	6.8	999	17	AAT42500 Mature HCPB coding
17	107.4	6.8	1053	17	AAT42497 Mature HCPB-(His)6
18	107.4	6.8	1263	17	AAT42506 Human pancreatic c
19	107.4	6.8	1284	17	AAT42506 Human pancreatic c
20	104.2	6.6	1053	20	AAV24804 Human carboxypepti
21	104.2	6.6	1059	17	AAT42511 Modified HCPB (D25
22	104.2	6.6	1059	17	AAT42512 Modified HCPB (D25
23	104.2	6.6	1059	18	AAT62787 Carboxypeptidase B
24	104.2	6.6	1059	18	AAT62788 Carboxypeptidase B
25	102.6	6.5	1059	18	AAT62805 Carboxypeptidase B
26	102.6	6.5	1059	18	AAT62806 Carboxypeptidase B
27	102.6	6.5	1059	18	AAT62791 Carboxypeptidase B
28	102.6	6.5	1059	18	AAT62801 Carboxypeptidase B
29	101	6.4	1059	18	AAT62789 Carboxypeptidase B
30	101	6.4	1059	18	AAT62790 Carboxypeptidase B
31	101	6.4	1870	18	AAV17322 DNA encoding human
32	101	6.4	1870	19	AAV17322 DNA encoding human
33	101	6.4	2154	18	AAV17331 Human carboxypepti
34	99.4	6.3	1059	18	AAT62808 Carboxypeptidase B
35	99.4	6.3	1059	18	AAT62809 Carboxypeptidase B
36	99.2	6.3	1059	18	AAT62800 Carboxypeptidase B
37	97.8	6.2	1059	18	AAT62807 Carboxypeptidase B
38	97.8	6.2	1059	18	AAT62803 Carboxypeptidase B
39	96.2	6.1	1059	18	AAT62802 Carboxypeptidase B
40	92.8	5.9	1311	22	AAD03837 Human carboxypepti
41	92.8	5.9	2128	22	AAH34774 Human colon cancer
42	91.6	5.8	936	22	AAH34774 Human breast cancer
43	88.6	5.6	527	22	AAL22383 Human carboxypepti
44	88.6	5.6	1050	22	AAD03838 Human carboxypepti
45	87.6	5.6	1200	22	AAD05477 Novel human protea

ALIGNMENTS

RESULT 1

AA081962 standard; cDNA; 1573 BP.

ID AAC81962;

AC AAC81962;

AC 01-MAR-2001 (first entry)

DE Human brain carboxypeptidase B cDNA.

DE Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase; treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine; cerbroprotective; antialzheimer's; neuroprotective;

KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;

KW Down's syndrome; head trauma; ss.

OS Homo sapiens.

Key Location/Qualifiers

CDS 18..1100

FT /*tag= a

FT /product= "carboxypeptidase B"

XX WO200066717-A1.

XX 09-NOV-2000.

XX 01-MAY-2000; 2000WO-JP02878.

XX 30-APR-1999; 99JP-0125169.

XX (MATS/) MATSUMOTO A.

XX Matsumoto A;
 XX WPI: 2000-687534/67.
 DR P-PSDB: AAB11457.
 XX Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 XX Claim 2b; Page 64-68; 84pp; Japanese.
 XX This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 XX Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;
 SQ

Query Match 100.0%; Score 1573; DB 21; Length 1573;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaaattgctgttggatgaagcttgcagccttgacgtctgtatcccatgttctct 60
 Db 1 agaaaattgctgttggatgaagcttgcagccttgacgtctgtatcccatgttctct 60
 QY 61 tctgtgagcagcagcttcttcgttcagagtgagcagagtcagtcgtctctctgaa 120
 Db 61 tctgtgagcagcagcttcttcgttcagagtgagcagagtcagtcgtctctctgaa 120
 QY 121 cctctagcagagtcagagcttctacagatcttctacacataagatgttctctgagc 180
 Db 121 cctctagcagagtcagagcttctacagatcttctacacataagatgttctctgagc 180
 QY 181 agccggttaacagcttgacatttctgaagaaaaaacaagtcatttttttgaatgcat 240
 Db 181 agccggttaacagcttgacatttctgaagaaaaaacaagtcatttttttgaatgcat 240
 QY 241 ctgagtgcagcaatgttgaagccatttaattgtgagcaggaattccatgacgttctctgc 300
 Db 241 ctgagtgcagcaatgttgaagccatttaattgtgagcaggaattccatgacgttctctgc 300
 QY 301 tggcagagcgtggaagatctatctcaacagagatcttccacagacagtcagcccccag 360
 Db 301 tggcagagcgtggaagatctatctcaacagagatcttccacagacagtcagcccccag 360
 QY 361 ccttcgcagctgactatgaacagatatacctaataatgaatatttcttggatagaat 420
 Db 361 ccttcgcagctgactatgaacagatatacctaataatgaatatttcttggatagaat 420
 QY 421 ttataacttgagagcagtcagatgttctcaaaaaatccacatggatcccatgttga 480
 Db 421 ttataacttgagagcagtcagatgttctcaaaaaatccacatggatcccatgttga 480
 QY 481 agtacccactctatgttttaaaagttctctgaaaagaaacaacagcacaatgcatat 540
 Db 481 agtacccactctatgttttaaaagttctctgaaaagaaacaacagcacaatgcatat 540
 QY 541 ggaattgactgtgaaatccatgacagagaatgagatctctctgtcttctgtgtgttca 600
 Db 541 ggaattgactgtgaaatccatgacagagaatgagatctctctgtcttctgtgtgttca 600
 QY 601 taggcataatcgaaatgtggaagaagaacggttcttctctatgtaacaatcatgtcatcg 660
 Db 601 taggcataatcgaaatgtggaagaagaacggttcttctctatgtaacaatcatgtcatcg 660

QY 661 gaacagacctgaatagcaacttcttccaaacactgtgtggaagaagtgcaccagtt 720
 Db 661 gaacagacctgaatagcaacttcttccaaacactgtgtggaagaagtgcaccagtt 720
 QY 721 cctcatgtcggaaacctactgtgacttatactctgagtcagaaacagaaatgaagcag 780
 Db 721 cctcatgtcggaaacctactgtgacttatactctgagtcagaaacagaaatgaagcag 780
 QY 781 tggctagtctcttgagaagaataatcaacagatataagcatatcatgcatcatat 840
 Db 781 tggctagtctcttgagaagaataatcaacagatataagcatatcatgcatcatat 840
 QY 841 actccagcagatagtggttccatattctctatacagaagaagaaagaagaccatagag 900
 Db 841 actccagcagatagtggttccatattctctatacagaagaagaaagaagaccatagag 900
 QY 901 aactgctctagtagcagagtgaaagcagcttgctgtattgtgacaacaactagtaaaatcca 960
 Db 901 aactgctctagtagcagagtgaaagcagcttgctgtattgtgacaacaactagtaaaatcca 960
 QY 961 ggtataacatgagcagatggtctcagaacactatacctagctcctgaggtgaggagatt 1020
 Db 961 ggtataacatgagcagatggtctcagaacactatacctagctcctgaggtgaggagatt 1020
 QY 1021 ggaatcatgatttgggcatcaaatatctggtttacataaaaccccgtagagaagcttt 1080
 Db 1021 ggaatcatgatttgggcatcaaatatctggtttacataaaaccccgtagagaagcttt 1080
 QY 1081 tgcgcgtgctctaaatagcttgcagtgcatataggaatgtttaatgagccctgatttta 1140
 Db 1081 tgcgcgtgctctaaatagcttgcagtgcatataggaatgtttaatgagccctgatttta 1140
 QY 1141 tcaattgcttcgttattttaattactgattccagcaagaacaatcatgtatcagat 1200
 Db 1141 tcaattgcttcgttattttaattactgattccagcaagaacaatcatgtatcagat 1200
 QY 1201 tattttaagtttatccgtaagtttgaataaagaatttccatctctctgttctgcag 1260
 Db 1201 tattttaagtttatccgtaagtttgaataaagaatttccatctctctgttctgcag 1260
 QY 1261 agaacctaataagtgctactcttgccatlaagcagagactaggttcacgtcttttaacct 1320
 Db 1261 agaacctaataagtgctactcttgccatlaagcagagactaggttcacgtcttttaacct 1320
 QY 1321 ttaaaaaaaattgttaaaagtctagttactacttcttcttctgtgatttcgagcttgact 1380
 Db 1321 ttaaaaaaaattgttaaaagtctagttactacttcttcttctgtgatttcgagcttgact 1380
 QY 1381 agccatctcaagcaacttctgacgcttgcagtcagcagatctcaagaagttaatacaagaat 1440
 Db 1381 agccatctcaagcaacttctgacgcttgcagtcagcagatctcaagaagttaatacaagaat 1440
 QY 1441 catctcagctgcatcatgtgactctactcaacaaaagaaggtgtgcagaagtatactta 1500
 Db 1441 catctcagctgcatcatgtgactctactcaacaaaagaaggtgtgcagaagtatactta 1500
 QY 1501 aagatctctgtccaaatttcaataaatttcttctctcttcttaaaaaaaaaaaaaaa 1560
 Db 1501 aagatctctgtccaaatttcaataaatttcttctctcttcttcttaaaaaaaaaaaaaaa 1560
 QY 1561 aaaaaaaaaaaaaa 1573
 Db 1561 aaaaaaaaaaaaaa 1573

RESULT 2
 AAT11671
 ID AAT11671 standard; DNA; 1749 BP.
 XX AAT11671;
 XX
 DT 12-APR-1996 (first entry)
 XX

Human plasma carboxypeptidase B coding sequence.

DE XX Plasma carboxypeptidase B; hpcpb; antibody; detection;
KW purification; plasminogen; affinity column; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 41..1312
FT /tag= a
FT /product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT /tag= b
FT mat_peptide 107..1309
FT /tag= c
XX US5474901-A.
XX 12-DEC-1995.
XX 01-FEB-1991; 91US-0649591.
XX 01-FEB-1991; 91US-0649591.
XX 14-OCT-1992; 92US-0959944.
XX 15-DEC-1993; 93US-0167727.
XX 19-JUL-1994; 94US-0277540.
XX (GENTECH) GENENTECH INC.
XX Drayna DT, Eaton DL;
XX WPI; 1996-039508/04.
XX P-PSDB; AAR90293.
XX Antibody to human plasma carboxypeptidase B - useful for detecting
PT and purifying hpcpb for use in treating clotting disorders e.g.
PT haemophilia A
XX Disclosure; Figure 4; 40pp; English.
XX An antibody which specifically binds human plasma carboxypeptidase B
CC (hpcpb) and does not cross react with other carboxypeptidases is
CC useful for the detection of hpcpb in vitro. The antibody is also
CC used for purifying hpcpb from a sample. Purification comprises
CC passing a sample thought to contain hpcpb over either a column to
CC which antibody has been bound, or a plasminogen affinity column,
CC eluting the column and then recovering the fraction containing the
CC hpcpb.
XX Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other:

Query Match 87.2%; Score 1372; DB 17; Length 1749;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

QY 1 agaaaattgctgttggaatgaagcttgcagccttgacgtcttaccattgttctct 60
DB 24 agaaaattgctgttggaatgaagcttgcagccttgacgtcttaccattgttctct 83
QY 61 tcttgagcagagatcttctgcgtccagagtgagcgaagtttaagtcttctccaaaga 120
DB 84 tcttgagcagagatcttctgcgtccagagtgagcgaagtttaagtcttctccaaaga 143
QY 121 cctctagcagaattcaagttctacagaatcttactacaacaatagatgttctctgcg 180
DB 144 cctctagcagaattcaagttctacagaatcttactacaacaatagatgttctctgcg 203
QY 161 agcggtaacagctgacattctatgtgaagaaacaaagtcatttttttgaatgcat 240
DB 204 agcggtaacagctgacattctatgtgaagaaacaaagtcatttttttgaatgcat 263
QY 241 ctgagtcgacaaatgtgaagcccatltaaatgtgagcggaattccatgcatgtcttcg 300

DB 264 ctgagtcgacaaatgtgaagcccatltaaatgtgagcggaattccatgcatgtcttcg 323
QY 301 tggcaacgttggaagattctatttcaacagaagattccacagacagtcagcccgag 360
DB 324 tggcagcgttggaagattctatttcaacagaagattccacagacagtcagcccgag 383
QY 361 cctcgcacgttactatgaacagtlcactcaactaaatgaatctcttggatgaat 420
DB 384 cctcgcacgttactatgaacagtlcactcaactaaatgaatctcttggatgaat 443
QY 421 ttataactgagaggcatctctgattatgcttaaaaaatccatttgatctcattgaga 480
DB 444 ttataactgagaggcatctctgattatgcttaaaaaatccatttgatctcattgaga 503
QY 481 agtacccactctagttttaagagttcttggaagaagaacaaacagcaaaaatgcatat 540
DB 504 agtacccactctagttttaagagttcttggaagaagaacaaacagcaaaaatgcatat 563
QY 541 ggaattgactgtggaatccatgccaagaatggaatcctctctcttctgtgtgttca 600
DB 564 ggaattgactgtggaatccatgccaagaatggaatcctctctcttctgtgtgttca 623
QY 601 taggcat----- 608
DB 624 taggcataactcaattctatggaataagggcaataataccaaatctcctgagcttg 683
QY 609 ----- 609
DB 684 tggattctcatgttatgcccgggtggttaagtgtgagcggttatgactctcatggaanaa 743
QY 610 atcgaatgtggaagaagaacggttcttctctatgccaacaatcatgcatcgagaagacc 669
DB 744 atcgaatgtggaagaagaacggttcttctctatgccaacaatcatgcatcgagaagacc 803
QY 670 tgaatgcaacttgcctccaaacactggtgtgaggaaggtgcatccagttcctatgct 729
DB 804 tgaatgcaacttgcctccaaacactggtgtgaggaaggtgcatccagttcctatgct 863
QY 730 cggaaacctactgtgacttaccctgagtcgaagcagaagtgagagcggtgagtg 789
DB 864 cggaaacctactgtgacttaccctgagtcgaagcagaagtgagagcggtgagtg 923
QY 790 tcttgagaagaatatcaaccagattaaagcatatcagcatgcatcactaccagc 849
DB 924 tcttgagaagaatatcaaccagattaaagcatatcagcatgcatcactaccagc 983
QY 850 atatagtgcttccatatcttccatacagaagtgaaagcagaacccatggagactgtcc 909
DB 984 atatagtgcttccatatcttccatacagaagtgaaagcagaacccatggagactgtcc 1043
QY 910 tagtaacagtggaagagttcgtctattgacaaaactgtaaaaatatacaggtataac 969
DB 1044 tagtaacagtggaagagttcgtctattgacaaaactgtaaaaatatacaggtataac 1103
QY 970 atggccatggtctcagaacattataccatgctctcgaagtgaggagatlgatcatg 1029
DB 1104 atggccatggtctcagaacattataccatgctctcgaagtgaggagatlgatcatg 1163
QY 1030 attgggcatcaaatatgt----- 1050
DB 1164 attgggcatcaaatatgt----- 1223
QY 1051 -----ttacaacaaacccactgtagagaacgtttgcccgtgtctcaaaa 1097
DB 1224 tgcgtccgagaggtatcacacaaacccactgtagagaacgtttgcccgtgtctcaaaa 1283
QY 1098 tagcttgcatgtcatataggaatglttaatgcccgtatltatcatctcgtccgtat 1157
DB 1284 tagcttgcatgtcatataggaatglttaatgcccgtatltatcatctcgtccgtat 1343
QY 1158 tttaatttcatgattccagaagaacaaatcatatgataagttatatttaagtttattc 1217


```

Db 804 tgaataggaacttctccaaacacgtgtgtaggaagtgatccacagttccatgct 863
Oy 730 cggaaacctactgttgacttattcctgagtcagaaccagaagtgaagcagtgtagt 789
Db 864 cggaaacctactgttgacttattcctgagtcagaaccagaagtgaagcagtgtagt 923
Oy 790 tcttgaagaagaatttcacacagatlaaagcatatcagcgtatctatctccacg 849
Db 924 tcttgaagaagaatttcacacagatlaaagcatatcagcgtatctatctccacg 983
Oy 850 ataatgtttccatattccctacacgaagtaaaagcaaacatctgagactgtcc 909
Db 964 ataatgtttccatattccctacacgaagtaaaagcaaacatctgagactgtcc 1043
Oy 910 tagtagcagagtagaagcagttcgtctattgtacaaactaglaaaaaatacaggtatcac 969
Db 1044 tagtagcagagtagaagcagttcgtctattgtagaaactaglaaaaaatacaggtatcac 1103
Oy 970 atggccatggtcagaacactatactcctgagtgagtgaggagatgagatctatg 1029
Db 1104 atggccatggtcagaacactatactcctgagtgagtgaggagatgagatctatg 1163
Oy 1030 atttgagcatcaaatattcgtt----- 1050
Db 1164 atttgagcatcaaatattcgtt----- 1223
Oy 1051 -----ttacatcaaacaccacgtgtagaagaagcttttgcgctgtctctaa 1097
Db 1224 tgcgtcggagcgttaccatcaaacaccacgtgtagaagaagcttttgcgctgtctctaa 1283
Oy 1098 tagcttgacatgcatcattaggaatgtttaagccctgattatcaatctgtctccgtat 1157
Db 1284 tagcttgacatgcatcattaggaatgtttaagccctgattatcaatctgtctccgtat 1343
Oy 1158 ttaattactactatcccgacgaacaaatcatgtatcagaattatatttaagttttac 1217
Db 1344 ttaattactactatcccgacgaacaaatcatgtatcagaattatatttaagttttac 1403
Oy 1218 cgtagtttgaataaagaatttcctatctctgtgtctgctgcagagaacctaagaagtc 1277
Db 1404 cgtagtttgaataaagaatttcctatctctgtgtctgctgcagagaacctaagaagtc 1463
Oy 1278 acttggcattaaagcagactagggttcatgtcttttaccctttaaaaaaattgttaa 1337
Db 1464 acttggcattaaagcagactagggttcatgtcttttaccctttaaaaaaattgttaa 1523
Oy 1338 aagctagttactacttttcttgatttgcagcttgtagcagcatctaaagaact 1397
Db 1524 aagctagttactacttttcttgatttgcagcttgtagcagcatctaaagaact 1583
Oy 1398 ttcgacgcttgtagcagcatctcaagcaagtttaatacaagaatctcaagctgtagat 1457
Db 1584 ttcgacgcttgtagcagcatctcaagcaagtttaatacaagaatctcaagctgtagat 1643
Oy 1458 tggatccctactcaacaaagaaggtgtgtcagaagtaataaagaattctgtctcaaa 1517
Db 1644 tggatccctactcaacaaagaaggtgtgtcagaagtaataaagaattctgtctcaaa 1703
Oy 1518 ttttcaataaattctctctcctttaaaaaaataaaaaa 1563
Db 1704 ttttcaataaattctctctcctttaaaaaaataaaaaa 1749

```

```

RESULT 4
AA041001
ID AA041001 standard; CDNA; 1749 BP.
XX
AC AA041001;
XX
XX 24-AUG-1993 (first entry)
XX
DE Human plasma carboxypeptidase B gene.
XX

```

```

KM PCPB: tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT sig_peptide 41..106
FT mat_peptide /tag= a
FT /tag= b
FT /product= PCPB
FT misc_feature 134..177
FT /tag= C
FT /function= probe
FT /note= "used to obtain full-length clones"
PD US5206161-A.
XX
XX 27-APR-1993.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI; 1993-151724/18.
XX
XX P-PSDB; AAR36273.
XX
XX
XX New human plasma carboxypeptidase B - used as haemostatic
XX regulator for clotting blood, partic. for treating blood clotting
XX disorders, e.g. haemophilia
XX
XX
XX PS Disclosure; Fig 4; 40pp; English.
XX
XX Human plasma carboxypeptidase B was isolated from human plasma and
XX partially sequenced. Oligonucleotide primers were designed based on
XX the partial amino acid sequences. The primers were used in a PCR
XX amplification to identify cDNA encoding PCPB from a human liver cDNA
XX library. The PCR product was capable of encoding the first 37 amino
XX acids of PCPB; a 46mer probe was used to obtain the full-length
XX sequence which, although disclosed in the specification, is not
XX claimed. PCPB inhibits the enzymatic conversion by tPA of
XX plasminogen to plasmin in the presence of fibrinogen.
XX
SQ Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

```

```

Query Match 87.1%; Score 1370.4; DB 14; Length 1749;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 6; Indels 163; Gaps 2;

```

```

Oy 1 agaaaattgctgttgatgaagaactttgacgcttgacgctctgttaccattgtctct 60
Db 24 agaaaattgctgttgatgaagaactttgacgcttgacgctctgttaccattgtctct 83
Oy 61 tctgtagcagcagctgtcttcgcttccagagtggtgccaagttcagtgctctccacaga 120
Db 84 tctgtagcagcagctgtcttcgcttccagagtggtgccaagttcagtgctctccacaga 143
Oy 121 cctctagcgaagttcaagttctacagaatcttactacaacatagagatgtctctggc 180
Db 144 cctctagcgaagttcaagttctacagaatcttactacaacatagagatgtctctggc 203
Oy 181 agccgtaacagctgacgtcttatgttgaagaagaaagcaagtttttttgaagcat 240
Db 204 agccgtaacagctgacgtcttatgttgaagaagaaagcaagtttttttgaagcat 263
Oy 241 ctgattcgacaatgtgaagccatttaaatgtgacggaattccatgagtgcttgc 300
Db 264 ctgattcgacaatgtgaagccatttaaatgtgacggaattccatgagtgcttgc 323

```

```

OY 301 tggcagacgtggaagatcttatacaacagcagattccaacgacacagtcagcccccag 360
    |||
Db 324 tggcagacgtggaagatcttatacaacagcagattccaacgacacagtcagcccccag 383
OY 361 ccttcgacgtgtaatagaagatatacctataaagaatataatcttcttggatagat 420
    |||
Db 384 ccttcgacgtgtaatagaagatatacctataaagaatataatcttcttggatagat 443
OY 421 ttaatactgaagcagatcctgatactgctatacaaaaatccaattgacccatttgaga 480
    |||
Db 444 ttaatactgaagcagatcctgatactgctatacaaaaatccaattgacccatttgaga 503
OY 481 agtaccacactatgctttaaagttcttggaagaaagacaaacagcaaaatgccaat 540
    |||
Db 504 agtaccacactatgctttaaagttcttggaagaaagacaaacagcaaaatgccaat 563
OY 541 ggaatgactgtggaatcgaatgcccagagaatgacatcccgcttctgcttggttca 600
    |||
Db 564 ggaatgactgtggaatcgaatgcccagagaatgacatcccgcttctgcttggttca 623
OY 601 taaggcat----- 608
Db 624 taaggcatatacctaattctatggaataagggcaatatataccaatctcctgagcttg 683
OY 609 -----a 609
Db 684 tgaattctatgtaatgcccgtggtaatggaacggtatgactactacatgaaagaa 743
OY 610 atcgaatgtggaagaaagaaacgcttcttctatatgcaaaatcatatgcatcggaaagac 669
    |||
Db 744 atcgaatgtggaagaaagaaacgcttcttctatatgcaaaatcatatgcatcggaaagac 803
OY 670 tgaatacgaactgtctccaacacactggtggaagaaagtgacatccagttctcctatgct 729
    |||
Db 804 tgaatacgaactgtctccaacacactggtggaagaaagtgacatccagttctcctatgct 863
OY 730 cggaaacctactgtggaacttatccctgagtgcaagaaacagaaagtgaagcgatgctagct 789
    |||
Db 864 cggaaacctactgtggaacttatccctgagtgcaagaaacagaaagtgaagcgatgctagct 923
OY 790 tcttgagaagaagaatatacacaacagatlaaagatacatatgacatcaccacgc 849
    |||
Db 924 tcttgagaagaagaatatacacaacagatlaaagatacatatgacatcaccacgc 983
OY 850 atatagtgttccatctcctatacagaaaglaaagcaaaagacatlgaggaaactgtctc 909
    |||
Db 984 atatagtgttccatctcctatacagaaaglaaagcaaaagacatlgaggaaactgtctc 1043
OY 910 taagtagcagtggaagcagcttctgctatgtgcaaaaactagtaaaaataccaggtatatac 969
    |||
Db 1044 taagtagcagtggaagcagcttctgctatgtgcaaaaactagtaaaaataccaggtatatac 1103
OY 970 atggccatggtcagaaacctatacactagctcctggaagtgagggaacgattggtctatg 1029
    |||
Db 1104 atggccatggtcagaaacctatacactagctcctggaagtgagggaacgattggtctatg 1163
OY 1030 atttgggcatacaaatatcg----- 1050
    |||
Db 1164 atttgggcatacaaatatcg----- 1223
OY 1051 -----ttatacaaaacccacctgtagaagaagcttttgcgcgtctcctataaa 1097
    |||
Db 1224 tgcgtccgagcggtatacaataaacccacctgtagaagaagcttttgcgcgtctcctataaa 1283
OY 1098 taagctgacatgcatatgaagaatgttaatgcccctatttatacatcctcctcgat 1157
    |||
Db 1284 taagctgacatgcatatgaagaatgttaatgcccctatttatacatcctcctcgat 1343
OY 1158 tttaattactgattccagaagaagcaaatcatgtatcagatattttaaagttatc 1217
    |||
Db 1344 tttaattactgattccagaagaagcaaatcatgtatcagatattttaaagttatc 1403
OY 1218 cgtagtttgaataaagaatttctcctatccttctggttctgtcagaagaacctaataagtgt 1277

```

```

Db 1404 cgtagtttgaataaagaatttctcctatccttctggtctgtcagaagaactataaagtct 1463
    |||
OY 1278 acttggcattaaagcaagactaaggtatcatgtctttaaaccctttaaanaaaatgttaa 1337
    |||
Db 1464 acttggcattaaagcaagactaaggtatcatgtctttaaaccctttaaanaaaatgttaa 1523
OY 1338 aagctaaactaacctacttcttcttctgatttctgaagttctgactagccatcccaagcaact 1397
    |||
Db 1524 aagctaaactaacctacttcttcttctgatttctgaagttctgactagccatcccaagcaact 1583
OY 1398 ttgacgcttgaactagccatcctcaagcaagtttaataaagatacatctcagcgtgatcat 1457
    |||
Db 1584 ttgacgcttgaactagccatcctcaagcaagtttaataaagatacatctcagcgtgatcat 1643
OY 1458 tggatcctactcaacaagaaggaaggtgtgtaagaagtacattaaagatttctgctccaaa 1517
    |||
Db 1644 tggatcctactcaacaagaaggaaggtgtgtaagaagtacattaaagatttctgctccaaa 1703
OY 1518 tttaataaatttcttctccttcttaaaaaaataaaaaaataa 1563
    |||
Db 1704 tttaataaatttcttctccttcttaaaaaaataaaaaaataa 1749

```

```

RESULT 5
AAV74302
ID AAV74302 standard; cDNA; 1272 BP.
XX
AC AAV74302;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
XX
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
XX
KW polymorphism detection; thrombotic disease; ds.
XX
OS Homo sapiens.
XX
FN M09855645-A1.
XX
PD 10-DEC-1998.
XX
PE 02-JUN-1998; 98WO-EP03244.
XX
PR 03-JUN-1997; 97US-0869057.
XX
PA (SCHD ) SCHERING AG.
XX
PI Morser MJ, Nagashima M;
XX
DR WPI: 1999-045800/04.
XX
P-PSDB: AAW92270.
XX
PT Detecting new polymorphism of human plasma carboxypeptidase B -
PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
PS Example 1; Page 24; 35pp; English.
XX
CC This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphs of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphs in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphs within the general
CC population compared with populations known to be at risk and establishing

```

CC	AAFI18005	standard; DNA; 1400 BP.
XX	AAFI18005;	
XX	AAFI18005;	
XX	14-MAR-2001	(first entry)
XX		
DE	Lung cancer associated polynucleotide sequence SEQ ID 24.	
XX		
XX	Human; lung cancer associated protein; neuroprotective; cytosolic;	
KW	cardiac; immunomodulatory; muscular active; vulnary;	
KW	gastrointestinal; nephrotropic; antiinfective; gynecological;	
KW	antibacterial; diagnosis; neural disorder; immune disorder; reproductive;	
KW	proliferative disorder; wound healing; infectious disease; ds.	
XX		
OS	Homo sapiens.	
PN	WO20005180-A2.	
XX		
PD	21-SEP-2000.	
XX		
PE	08-MAR-2000; 2000WO-US05918.	
XX		
PR	12-MAR-1999; 99US-0124270.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
PA	(ROSE/) ROSEN C A.	
XX		
PI	Ruben SM;	
XX		
DR	WPI: 2000-587514/55.	
DR	P-PSDB: AAB58129.	
XX		
PT	Lung cancer associated gene sequences, referred to as lung cancer	
PT	antigens, useful for treatment, prevention, and diagnosis of disorders	
PT	such as lung cancer -	
XX		
PS	Claim 1; Page 507; 1425pp; English.	
XX		
XX	Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer	
CC	associated proteins represented in AAB58106 - AAB58548. Lung cancer	
CC	associated proteins and polynucleotide sequences, their agonists, and	
CC	antagonists may have neuroprotective; cytosolic; cardioactive;	
CC	immunomodulatory; muscular active general; vulnary; gastrointestinal	
CC	general; nephrotropic; antiinfective; gynecological; or antibacterial	
CC	activity. The invention also includes antibodies specific for the	
CC	protein or polynucleotide sequences. The lung cancer associated	

CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.

XX
 XX
 SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;

Query Match 54.3%; Score 853.8; DB 21; Length 1400;
 Best local Similarity 93.3%; Pred. No. 6.1e-196;
 Matches 934; Conservative 1; Mismatches 13; Indels 53; Gaps 2;

QY 607 ataatacgaatgtagaagaagaacggtcttctatgcgaacaatcattgcacggaacag 666
 DB 177 agaatcgaatctgtagaagaagaacggtcttctatgcgaacaatcattgcacggaacag 236
 QY 667 acctgaatacgaatctgtagaagaagaacggtcttctatgcgaacaatcattgcacggaacag 726
 DB 237 acctgaatacgaatctgtagaagaagaacggtcttctatgcgaacaatcattgcacggaacag 296
 QY 727 gctcgaagaactactgtagaagaagaacggtcttctatgcgaacaatcattgcacggaacag 786
 DB 297 gctcgaagaactactgtagaagaagaacggtcttctatgcgaacaatcattgcacggaacag 356
 QY 787 gttctctgagaagaagaatacaacagatgtaagcacaatcattgcacggaacag 846
 DB 357 gttctctgagaagaagaatacaacagatgtaagcacaatcattgcacggaacag 416
 QY 847 agcatatagtttccatctcctatcacaagaagtaagaagaacagatgtaggaggaactgtc 906
 DB 417 agcatatagtttccatctcctatcacaagaagtaagaagaacagatgtaggaggaactgtc 476
 QY 907 cctcagtagcagcagtagaagaagaacggtcttctatgcgaacaatcattgcacggaacag 966
 DB 477 cctcagtagcagcagtagaagaagaacggtcttctatgcgaacaatcattgcacggaacag 536
 QY 967 cacatgagcctgtagcagaagaacggtcttctatgcgaacaatcattgcacggaacag 1026
 DB 537 cacatgagcctgtagcagaagaacggtcttctatgcgaacaatcattgcacggaacag 596
 QY 1027 atgattggcagcaacaatctgtt----- 1050
 DB 597 atgattggcagcaacaatctgtt----- 656
 QY 1051 -----ttacatcaaacacacgtagaagaagctttgacggtcgtccta 1094
 DB 657 tcttgctgcgagcagcgttacaacaacacgtagaagaagctttgacggtcgtccta 716
 QY 1095 aatatagcttgacgtagcagtagaagaagctttgacggtcgtccta 1154
 DB 717 aatatagcttgacgtagcagtagaagaagctttgacggtcgtccta 776
 QY 1155 taatttaattactgtagcagtagaagaagctttgacggtcgtccta 1214
 DB 777 taatttaattactgtagcagtagaagaagctttgacggtcgtccta 836
 QY 1215 atccgtagtttgcataaagaatttccattccttggtctgtcagagaacctaataagt 1274
 DB 837 atccgtagtttgcataaagaatttccattccttggtctgtcagagaacctaataagt 896
 QY 1275 gctactttgacatgaagcagcagtaggtagcagcttttaccctttaaaaaaattg 1334
 DB 897 gctactttgacatgaagcagcagtaggtagcagcttttaccctt-aaaaaaattg 955
 QY 1335 taaaagctcgtacacacatttcttgcatttgcagcgtttgacgacatcccaagca 1394
 DB 956 taaaagctcgtacacacatttcttgcatttgcagcgtttgacgacatcccaagca 1015

QY 1395 acttcgacggttgcacatccacgaagaagtttaatacaaatcattcgcacgtagat 1454
 DB 1016 acttcgacggttgcacatccacgaagaagtttaatacaaatcattcgcacgtagat 1075
 QY 1455 catgtagcctcctacacaaagaaggtgtagaagaagtagcattaaagattctgtcc 1514
 DB 1076 catgtagcctcctacacaaagaaggtgtagaagaagtagcattaaagattctgtcc 1135
 QY 1515 aaatttcaataattctctctcctttaaataaaaaa 1555
 DB 1136 aaatttcaataattctctctcctttaaataaaaaa 1176

RESULT 7

AAH85957
 ID AAH85957 standard; DNA; 333 BP.

XX AAH85957;
 AC
 XX
 DT 27-FEB-2002 (first entry)

XX Human single nucleotide polymorphism containing DNA sequence #814.

DE Allelic marker: polymorphism; human; disease; diagnosis; treatment;
 KW phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;
 KW transgenic; single nucleotide polymorphism; SNP; ss.

OS Homo sapiens.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Location/Qualifiers
 replace(274,C,G)
 /*tag= a
 /standard_name= "single nucleotide polymorphism"

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Claim 1; Page 117; 330pp; English.
 PS This invention describes novel human nucleic acid segments (I)
 CC containing polymorphic sites. The polynucleotides of (I) are used for,
 CC e.g. correlating disease polymorphisms (or disease susceptibility) or
 CC other phenotypic traits (e.g. baldness, obesity, fertility, strength,
 CC response to drugs etc.); diagnosing and monitoring e.g. cancer,
 CC inflammation, heart or central nervous system diseases; detecting
 CC susceptibility to microbial infection; treating or preventing such
 CC diseases; forensic analysis; gene therapy; paternity testing; mapping
 CC genomic loci associated with phenotypic traits (and subsequent cloning
 CC of the genes responsible); and the production of transgenic organisms.
 CC Antibodies raised against (I) are useful as diagnostic and therapeutic
 CC tools and in drug screening. AAH85144 - AAH87644 represent the human
 CC DNA sequences containing diallelic polymorphic sites described in the
 CC invention.
 SQ Sequence 333 BP; 90 A; 64 C; 53 G; 117 T; 9 other;

	Query Match	20.6%	Score 324;	DB 20;	Length 333;
	Best Local Similarity	97.3%;	Pred. No. 1.7e-68;		
	Matches 324; Conservative	0;	Mismatches 9;	Indels	0; Gaps
Oy	1159 ttaattactcgtatcccaagaagccaataatcatgfatcatgaatattttaagtttatccc				1218
Dd	1 ttaatttaccgtatcccaagaagccaataatcatgfatcatgaatattttaagtttatccc				60
Oy	1219 gtagtttgataaaaagatcttcctatctcctgtgtctgcacagaacctaaatggtcta				1278
Dd	61 gtagtttgataaaaaagatcttcctatctcctgtgtctgcacagaacctaaatggtcta				120
Oy	1279 cttygcacataaggcagacgtagggtcatgtctcttttaaccccccttaaaaaaaaaatgtaaa				1338
Dd	121 ctttgccatttaaggagactaaggtcatgtctcttttaacccccctnnmmnmnttgtaa				180
Oy	1339 agtcagttaactactttttcttttgtattttcgagctttggacatgccatctcaagaacct				1398
Dd	181 agtcagttaactactttttcttttgtattttcgagctttggacatgccatctcaagaacct				240
Oy	1399 tcgaagcttgatcacatccatccaagaagatttaaccaaagatcatctcagctgatcat				1458
Dd	241 tcgaagcttgatcacatccatccaagaagatttaaccaaagatcatctcagctgatcat				300
Oy	1459 ggatcctactcaacaagaagggtggtcaga				1491
Dd	301 ggatcctactcaacaagaagggtggtcaga				333
RESULT	8				
AHAH87390	standard; DNA; 333 BP.				
XX	AAH87390;				
AC					
XX					
DT	27-FEB-2002 (first entry)				
Dd					
XX					
DE	Human single nucleotide polymorphism containing DNA sequence #2247.				
XX					
KW	Biallelic marker: polymorphism: human: disease: diagnosis: treatment;				
KM	phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;				
KM	transgenic; single nucleotide polymorphism; SNP; ss.				
OS	Homo sapiens.				
XX					
XX					
FH	Key location/Qualifiers				
FT	Variation replace(238,G)				
FT	/tag= a				
FT	/standard_name= "single nucleotide polymorphism"				
PN	MO9953095-A2.				
XX					
PD	21-OCT-1999.				
XX					
PF	30-MAR-1999; 99WO-US06893.				
XX					
PR	09-APR-1998; 98US-0057871.				
XX					
PA	(WHED) WHITEHEAD INST BIOMEDICAL RES.				
PI	Lander ES, Wang D, Hudson T;				
XX					
DR	WPI; 1999-620443/53.				
XX					
PT	This invention describes novel human nucleic acid segments (I)				
PT	containing polymorphic sites. The polynucleotides of (I) are used for,				
PT	e.g. correlating disease polymorphisms (or disease susceptibility) or				
PT	of disease -				
XX					
PS	Claim 1; Page 283; 330pp; English.				
XX					

CC	other phenotypic traits (e.g. baldness, obesity, fertility, strength,
CC	response to drugs etc.); diagnosing and monitoring e.g. cancer,
CC	inflammation, heart or central nervous system diseases; detecting
CC	susceptibility to microbial infection; treating or preventing such
CC	diseases; forensic analysis; gene therapy; paternity testing; mapping
CC	genomic loci associated with phenotypic traits (and subsequent cloning
CC	of the genes responsible); and the production of transgenic organisms
CC	Antibodies raised against (1) are useful as diagnostic and therapeutic
CC	tools and in drug screening. AAH85144 - AAH87644 represent the human
CC	DNA sequences containing biallelic polymorphic sites described in the
CC	invention.
XX	
SQ	Sequence 333 BP; 90 A; 64 C; 53 G; 117 T; 9 other;
Query Match 20 64; Score 324; DB 20; Length 333;	
Best Local Similarity 97.38; Pred. No. 1,7e-68;	
Matches 324; Conservative 0; Mismatches 9; Indels 0; Gaps	
QY	1159 ttaattctatcttcgcgcgaagcaaccaatcatctgtacagattattttaagttatcc 121
DB	1 ttaattctatcttcgcgcgaagcaaccaatcatctgtacagattattttaagttatcc 60
QY	1219 gtagctttgataaaagatttccctatccctgtgtctgtcgaagaaactaataagtgc 127
DB	61 gtagctttgataaaagatttccctatccctgtgtctgtcgaagaaactaataagtgc 120
QY	1279 cttagcatttaagcagaagctagggttaatgctcttttaccctttaaaaaattgtaaa 133
DB	121 cttagcatttaagcagaagctagggttaatgctcttttaccctttmmmmnnmtgtaaa 180
QY	1339 agctcagttacctactcttttcttggatttcgcagcttgcagtcacatcgaagaacct 139
DB	181 agctcagttacctactcttttcttggatttcgcagcttgcagtcacatcgaagaacct 240
QY	1399 tcgacgcttgcactgcagcatctcaagcaagtttaatacaagaatcatctcagctgacatt 145
DB	241 tcgacgcttgcactgcagcatctcaagcaagtttaatacaagaatcatctcagctgacatt 300
QY	1459 ggaatccactactcaacaaaaggaaggtgtgcaga 1491
DB	301 ggaatccactactcaacaaaaggaaggtgtgcaga 333
RESULT 9	
AAx11374	
ID	AAx11374 standard; DNA: 251 BP.
XX	
AC	AAx11374;
XX	
DT	30-MAR-1999 (first entry)
XX	
DE	Human biallelic polymorphic DNA fragment WI-7773b.
XX	
KW	Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW	detection; phenotypic typing; characteristic; infection; hereditary;
KW	autoimmune disease; cancer; inflammation; drug; therapy; medication;
KW	treatment; marker; SS.
XX	
OS	Homo sapiens.
XX	
PN	W09820165-A2.
XX	
PD	14-MAY-1998.
XX	
PF	05-NOV-1997; 97WO-US20313.
XX	
PR	06-NOV-1996; 96US-0030455.
XX	
EA	(WHED) WHITEHEAD INST BIOMEDICAL RES.
XX	
PI	Hudson T, Lander ES, Wang D;
XX	

DR WPI: 1998-286974/25.

XX New isolated nucleic acid segments from the human genome - used for
 PT determining polymorphic forms for use in e.g. forensics, paternity
 PT testing or phenotypic typing for disease

XX
 PS Claim 1; Page 152; 310pp; English.

CC AAX10269-X12937 are human DNA fragments which contain biallelic
 CC polymorphic markers which have been isolated using the primers
 CC represented in AAX09121-X10268. The base occupying the polymorphic site
 CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
 CC can be used in methods for determining polymorphic forms in an individual
 CC for use in e.g. forensics, paternity testing or for phenotypic typing for
 CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan
 CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
 CC familial hypercholesterolemia, polycystic kidney disease, hereditary
 CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous
 CC system, infection by pathogenic microorganisms, and characteristics such
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
 CC endurance, fertility, and susceptibility or receptivity to particular
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
 CC segments can also be used to produce medicaments for the treatment or
 CC prophylaxis of such diseases.

XX
 SQ Sequence 251 BP; 62 A; 44 C; 37 G; 98 T; 10 other;

Query Match 15.4%; Score 241.6; DB 19; Length 251;
 Best Local Similarity 96.0%; Pred. No. 1.1e-48;
 Matches 241; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 1159 ttaattcagatccgacgaacaaatcatgtatcagatatttttaagtatttc 1218
 DB 1 ttaattcagatccgacgaacaaatcatgtatcagatatttttaagtatttc 60

QY 1219 gtatgttgaataaagatttcctatccttggtctgtcagagaacctaaagtgtca 1278
 DB 61 gtatgttgaataaagatttcctatccttggtctgtcagagaacctaaagtgtca 120

QY 1279 cttgcataaaggacagagggttcattcttttaaccttaaaaaaatgtgaaa 1338
 DB 121 cttgcataaaggacagagggttcattcttttaaccttaaaaaaatgtgaaa 180

QY 1339 agtctaagtaacctttcttcttgatttcgacggttgactagacatccaagcaact 1398
 DB 181 agtctaagtaacctttcttcttgatttcgacggttgactagacatccaagcaact 240

QY 1399 tcgaagtttga 1409
 DB 241 tcgaagtttga 251

RESULT 10
 AAT35760
 ID AAT35760 standard; cDNA: 927 BP.
 AC AAT35760;
 XX
 DT 09-NOV-1996 (first entry)
 XX
 DE Rat mature carboxypeptidase B cDNA.
 XX
 KW Carboxypeptidase-B; pro-enzyme; protease; insulin; ss.
 OS Rattus sp.
 XX
 FH Key Location/Qualifiers
 FT primer_bind complement(1..50)
 FT /tag- a

FT primer_bind /note= "Mature CPB 5'-end primer"
 FT 898..927
 FT /*tag- b
 FT /note= "CPB 3' end primer"

XX
 PN W09623064-A1.
 XX
 PD 01-AUG-1996.
 XX
 PF 25-JAN-1996; 96WO-US00995.
 XX
 PR 25-JAN-1995; 95US-0378233.
 XX
 PA (BIOT-) BIOTECHNOLOGY GEN CORP.
 XX
 PI Fulga N, Gorecki M, Hartman J, Mendelovitch S;
 DR WPI: 1996-362688/36.
 DR P-PSDB; AAM00602.

XX
 PT Purified active recombinant carboxypeptidase B prodn. - by
 PT expressing DNA encoding the pro-enzyme, folding and enzymatic
 PT cleavage to give active form, used e.g. for insulin prodn.

PS Example 1; Page 37-38; 49pp; English.

XX
 CC A cDNA sequence (AAT35760) codes for the mature form (AAM00602) of
 CC rat carboxypeptidase B (CPB). A full-length sequence including
 CC cDNA (AAT35759) coding for the activation peptide (AAM00601) can
 CC be obtained by PCR amplification (see also AAT35756 and AAT35758) of cDNA
 CC derived from Sprague-Dawley rat pancreas. Pro-CPB can be expressed
 CC in host cells, e.g. Escherichia coli, and subsequently recovered,
 CC refolded and cleaved with trypsin to yield the active enzyme. CPB
 CC produced this way is cheaper than porcine pancreatic enzyme, and is
 CC free of other proteases.

XX
 SQ Sequence 927 BP; 253 A; 226 C; 213 G; 235 T; 0 other;

Query Match 7.9%; Score 124; DB 17; Length 927;
 Best Local Similarity 57.9%; Pred. No. 3.6e-20;
 Matches 278; Conservative 0; Mismatches 195; Indels 7; Gaps 3;

QY 613 gaatgtgagaagaacgcttcttctatcgaacaatcatgtatcagacagactga 672
 DB 362 gaatgtgagaagaacgcttcttctatcgaatgttctctgtgtagtaccacca 421

QY 673 atagcaacttgctcccaaacactggtgtgagaaggtgcatcagttcctatgctcg 732
 DB 422 atagcaacttgctcccaaacactggtgtgagaaggtgcatcagttcctatgctcg 478

QY 733 aaactactgttgacatttctcctgagtcagacaagaagtgaagcgatgtcttct 792
 DB 479 aaactactgttgacacgcccagagctctgaaaaagagaagaagcgccctgcatattca 538

QY 793 tgagaagaatcaacaacagataaagcatatcatcagatgcatltaactccacagata 852
 DB 539 tccgacaacaacctccaccatcaagaagcctcagaccatccactactacacagatga 598

QY 853 tagtgttcataattcctatcaacgagtaagaagcaaaagccttgaggaactgtcctag 912
 DB 599 tgcctcacccttactcctactgactacaacactgctctgagaactaggaattgagccc 658

QY 913 taagcagltgaagaagltcgtcgtatagacaacaaactgataaataacaggtatcacatg 972
 DB 659 tggltgaaggtcgcgcaagaagactgtgc---actcgtcatgacccaagaatatacatg 715

QY 973 gccatggtctagaacacctatacctagctctcgtgaggtggtgagcatgtgatcatgatt 1032
 DB 716 gccacagagctacaacaatcatcctcgtctgtgggagatcgtacgagctgtctatgatac 775

QY 1033 tgggcataaatatcttgttaacataaacaccacctgtgagaagccttttgcgcgtcttc 1092


```

FT Variation      replace(88,C)
FT               /*tag= a
FT               /standard_name= "single nucleotide polymorphism"
XX
XX PN WO953095-A2.
XX FT
XX PD 21-OCT-1999.
XX PF
XX PR 30-MAR-1999; 99WO-US06893.
XX PR 09-APR-1998; 98US-0057871.
XX PA (WHEHD ) WHITEHEAD INST BIOMEDICAL RES.
XX PI Lander ES, Wang D, Hudson T;
XX DR WPI: 1999-620443/53.
XX
XX PT Polymorphic human genomic sequences and related allele-specific probes
XX PT and primers, useful for genetic analysis, e.g. diagnosis and monitoring
XX PT of disease
XX PS Claim 1; Page 91; 330pp; English.
XX
CC This invention describes novel human nucleic acid segments (I)
CC containing polymorphic sites. The polynucleotides of (I) are used for,
CC e.g. correlating disease polymorphisms (or disease susceptibility) or
CC other phenotypic traits (e.g. baldness, obesity, fertility, strength,
CC response to drugs etc.); diagnosing and monitoring e.g. cancer,
CC inflammation, heart or central nervous system diseases; detecting
CC susceptibility to microbial infection; treating or preventing such
CC diseases; forensic analysis; gene therapy; paternity testing; mapping
CC genomic loci associated with phenotypic traits (and subsequent cloning
CC of the genes responsible); and the production of transgenic organisms.
CC Antibodies raised against (I) are useful as diagnostic and therapeutic
CC tools and in drug screening. AAH85144 - AAH87644 represent the human
CC DNA sequences containing biallelic polymorphic sites described in the
CC invention.
XX
XX SQ Sequence 121 BP; 33 A; 22 C; 29 G; 36 T; 1 other;

```

```

XX
XX Key      Location/Qualifiers
XX CDS      1..1248
XX          /*tag= a
XX          /product= "pancreatic carboxypeptidase B"
XX          /note= "No start codon given"
XX
XX PN WO9835988-A1.
XX PD
XX PR 20-AUG-1998.
XX PR 10-FEB-1998; 98WO-GB00415.
XX PR 29-OCT-1997; 97GB-0022727.
XX PR 14-FEB-1997; 97GB-0003104.
XX PR 18-OCT-1997; 97GB-0022003.
XX
XX PA (ZENE ) ZENEPA (ZENE ) ZENECA LTD.
XX PI Edge MD;
XX
XX DR WPI: 1998-467168/40.
XX DR P-PSDB; AAM74476.
XX
XX PT New modified pro-domain of carboxy-peptidase B - enhances expression
XX PT of co-expressed proteins for production of recombinant
XX PT carboxy-peptidase or its fusions with antibodies, used, e.g. in
XX PT enzyme prodnrg therapy
XX
XX PS Example 1; Page 52-53; 83pp; English.
XX
XX CC The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
XX CC modified pro-domain of CPB on a separate gene to enhance recombinant
XX CC expression. This process can be used to produce recombinant CPB in
XX CC eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
XX CC insulin production and protein sequencing, while its fusions with
XX CC antibodies are useful in antibody-directed enzyme prodnrg therapy. The
XX CC modified pro-domain provide increased yields of recombinant CPB, possibly
XX CC by protecting the C-terminus against enzymatic degradation or by
XX CC increasing intracellular trafficking.
XX
XX SQ Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;

```

```

Query Match      7.6%; Score 120; DB 20; Length 121;
Best Local Similarity 99.2%; Pred. No. 1.7e-19;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Query Match      6.9%; Score 109; DB 19; Length 1263;
Best Local Similarity 54.4%; Pred. No. 1.6e-16;
Matches 264; Conservative 0; Mismatches 215; Indels 6; Gaps 2;

```

Db 1002 agaacctggcc---tcactgcacgacccaagtlacacataatgcccggagctacacacat 1058
Qy 992 ataccctagctccctggaggtggggaacgattggtatgatgattgggaatcaataatcgtt 1051
Db 1059 ctaccctgctctggtgggctctgaacgactggcttatgaccaaggaaatcagatatctcct 1118
Qy 1052 tacat 1056
Db 1119 cacct 1123

Search completed: September 18, 2002, 16:35:58
Job time: 4316 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2002, 15:25:32 ; Search time 54.5 Seconds

(without alignments)
7089.569 Million cell updates/sec

Title: US-09-980-881-1

Perfect score: 1573

Sequence: 1 agaaaatgctgttggatg.....aaaaaaaaaaaaaaaa 1573

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCrUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1372	87.2	1749	1 US-07-649-591B-2	Sequence 2, Appl
2	1372	87.2	1749	1 US-08-277-540-2	Sequence 2, Appl
3	1372	87.2	1749	1 US-08-430-787A-2	Sequence 2, Appl
4	918	58.4	1272	2 US-08-869-057-1	Sequence 1, Appl
5	124	7.9	927	2 US-08-782-760-5	Sequence 5, Appl
6	124	7.9	927	5 PCr-US96-00995-5	Sequence 5, Appl
7	123.4	7.8	921	1 US-08-696-139-3	Sequence 3, Appl
8	123.4	7.8	1215	1 US-08-696-139-1	Sequence 1, Appl
9	107.4	6.8	999	2 US-08-860-882A-67	Sequence 67, Appl
10	107.4	6.8	1053	2 US-08-860-882A-64	Sequence 64, Appl
11	107.4	6.8	1263	2 US-08-860-882A-56	Sequence 56, Appl
12	107.4	6.8	1284	2 US-08-860-882A-71	Sequence 71, Appl
13	104.2	6.6	1059	2 US-08-860-882A-74	Sequence 74, Appl
14	104.2	6.6	1059	2 US-08-860-882A-77	Sequence 77, Appl
15	101	6.4	1870	4 US-09-171-945-112	Sequence 112, App
16	101	6.4	2154	4 US-09-171-945-124	Sequence 124, App
17	71.6	4.6	1251	4 US-08-640-906-3	Sequence 3, Appl
18	70.6	4.5	1251	4 US-09-395-936-3	Sequence 3, Appl
19	70.6	4.5	1257	4 US-08-640-906-1	Sequence 1, Appl
20	70.6	4.5	1257	4 US-09-395-936-1	Sequence 1, Appl
21	46.8	3.0	285	2 US-08-782-760-2	Sequence 2, Appl
22	46.8	3.0	285	5 PCr-US96-00995-2	Sequence 2, Appl
23	45.6	2.9	8920	4 US-08-446-855A-1	Sequence 1, Appl
24	45.6	2.9	8920	4 US-09-150-741-1	Sequence 1, Appl
25	45	2.9	4665	4 US-08-948-378A-7	Sequence 7, Appl
26	45	2.9	4665	4 US-09-169-425C-7	Sequence 7, Appl
27	44.8	2.8	7218	1 US-08-232-463-14	Sequence 14, Appl

28	43.8	2.8	880	1 US-08-616-368A-7	Sequence 7, Appl
29	43.8	2.8	880	3 US-09-054-298-7	Sequence 7, Appl
30	43.8	2.8	880	4 US-08-818-655-7	Sequence 7, Appl
31	43.8	2.8	1610	4 US-09-013-896A-1	Sequence 1, Appl
32	43.2	2.7	9521	4 US-08-972-218-2	Sequence 2, Appl
33	43.2	2.7	13905	4 US-08-972-218-1	Sequence 1, Appl
34	42.4	2.7	1129	4 US-09-227-357-40	Sequence 40, Appl
35	42.4	2.7	1472	4 US-08-781-420-10	Sequence 10, Appl
36	42.4	2.7	1472	4 US-08-781-420-12	Sequence 12, Appl
37	42.4	2.7	1472	4 US-08-874-102-10	Sequence 10, Appl
38	42.4	2.7	1472	4 US-08-874-102-12	Sequence 12, Appl
39	42.4	2.7	1681	4 US-08-874-102-46	Sequence 46, Appl
40	42.4	2.7	1681	4 US-08-874-102-48	Sequence 48, Appl
41	42	2.7	2158	1 US-08-698-551-1	Sequence 1, Appl
42	42	2.7	2158	1 US-08-602-228-1	Sequence 1, Appl
43	42	2.7	2158	2 US-08-649-341A-1	Sequence 1, Appl
44	42	2.7	2158	2 US-08-494-440B-1	Sequence 1, Appl
45	42	2.7	2158	2 US-08-533-901B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-07-649-591B-2
; Sequence 2, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: hybridization probe
; LOCATION: 133 to 178
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: potential clip site
; LOCATION: 380 to 382
; IDENTIFICATION METHOD:

OTHER INFORMATION:
 FEATURE: signal sequence
 NAME/KEY: 41 to 106
 LOCATION: IDENTIFICATION METHOD:
 OTHER INFORMATION:
 US-07-649-591B-2

Query Match 87.2%; Score 1372; DB 1; Length 1749;
 Best Local Similarity 90.3%; Pred. No. 0;
 Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

1 agaaaattgctgttggatgaagcttgcagccttgagcttgccttgaccattgtctct 60
 24 aaaaaattgctgttggatgaagcttgcagccttgagcttgccttgaccattgtctct 83
 61 tctgtgaagcagatgtcttcgcttcagagcttgcagccttgagcttgccttgaccattgtctct 120
 84 tctgtgaagcagatgtcttcgcttcagagcttgcagccttgagcttgccttgaccattgtctct 143
 121 cctctgaagcagatgtcttcgcttcagagcttgcagccttgagcttgccttgaccattgtctct 180
 144 cctctgaagcagatgtcttcgcttcagagcttgcagccttgagcttgccttgaccattgtctct 203
 181 agcggtaacagctgacaccttatttgagagaaaaaagaagctcatttttctgaatgat 240
 204 acccggttaacagctgacaccttatttgagagaaaaaagaagctcatttttctgaatgat 263
 241 ctgaatctgaacagctgacaccttatttgagagaaaaaagaagctcatttttctgaatgat 300
 264 ctgaatctgaacagctgacaccttatttgagagaaaaaagaagctcatttttctgaatgat 323
 301 tggagacgctggaagacaccttatttgagagaaaaaagaagctcatttttctgaatgat 360
 324 tggagacgctggaagacaccttatttgagagaaaaaagaagctcatttttctgaatgat 383
 361 cctcgcagctggaagacaccttatttgagagaaaaaagaagctcatttttctgaatgat 420
 384 cctcgcagctggaagacaccttatttgagagaaaaaagaagctcatttttctgaatgat 443
 421 ttaatactgaagacgacaccttatttgagagaaaaaagaagctcatttttctgaatgat 480
 444 ttaatactgaagacgacaccttatttgagagaaaaaagaagctcatttttctgaatgat 503
 481 agtacccactctatgtttttaaaggttcttgagagaaaaaagaagctcatttttctgaatgat 540
 504 agtacccactctatgtttttaaaggttcttgagagaaaaaagaagctcatttttctgaatgat 563
 541 ggaatgactgtggaatgcacatgcagagaaatgagcttctcctgcttctgcttgagcttca 600
 564 ggaatgactgtggaatgcacatgcagagaaatgagcttctcctgcttctgcttgagcttca 623
 601 taggacat----- 608
 624 taggacat----- 683
 609 -----a 609
 684 tggatttctatgttatgcccgtgtttatgtgacggttatgactactatgaaataa 743
 610 atcgaatgtggaagaaagacgcttcttctatgagaaatcattgcatggaagaccc 669
 744 atcgaatgtggaagaaagacgcttcttctatgagaaatcattgcatggaagaccc 803
 670 tgaatgaacactgtctcacaacactgtgtgagaaagtgatccagcttctcctatgct 729
 804 tgaatgaacactgtctcacaacactgtgtgagaaagtgatccagcttctcctatgct 863
 730 cggaaacactgtgtgacttaccctgaagtcagaaacgaagtgaaagcagtgagctatgt 789
 864 cggaaacactgtgtgacttaccctgaagtcagaaacgaagtgaaagcagtgagctatgt 923

790 tcttgaagaaataatcaaccagatgaagcatalcagcagctatcatalcaccagc 849
 924 tcttgaagaaataatcaaccagatgaagcatalcagcagctatcatalcaccagc 983
 850 atatagtgttccatattcctatcagagaaatgaagaaacatgaagaaatgtctc 909
 984 atatagtgttccatattcctatcagagaaatgaagaaacatgaagaaatgtctc 1043
 910 tagtgcagatgaagcagctgtgctatgtgagaaatgaagaaatgaagaaatgtctc 969
 1044 tagtgcagatgaagcagctgtgctatgtgagaaatgaagaaatgaagaaatgtctc 1103
 970 atgagcagatgtcagaaacacttatacctagcctcctgagagtgaggagcattgacatg 1029
 1104 atgagcagatgtcagaaacacttatacctagcctcctgagagtgaggagcattgacatg 1163
 1030 atttgagcagaaatattcgt----- 1050
 1164 atttgagcagaaatattcgt----- 1223
 1051 -----ttacatcaaccacctgtgagaaagcttgcgctgtctctaa 1097
 1224 tgcctccgagcagctgtacatcaaccacctgtgagaaagcttgcgctgtctctaa 1283
 1098 tagcttgagcagatgaagaaatgaagaaatgaagaaatgaagaaatgaagaaatgaag 1157
 1284 tagcttgagcagatgaagaaatgaagaaatgaagaaatgaagaaatgaagaaatgaag 1343
 1158 ttaatactgaagacgacaccttatttgagagaaaaaagaagctcatttttctgaatgat 1217
 1344 ttaatactgaagacgacaccttatttgagagaaaaaagaagctcatttttctgaatgat 1403
 1218 cgtatgttgaataaagatttccatcctgtgtgtcgaagaaacctaagctat 1277
 1404 cgtatgttgaataaagatttccatcctgtgtgtcgaagaaacctaagctat 1463
 1278 acttgacatgaagcagactagagtgatcatttaccctttaaataaataatgttaa 1337
 1464 acttgacatgaagcagactagagtgatcatttaccctttaaataaataatgttaa 1523
 1338 aagctgaatgaagcagactagagtgatcatttaccctttaaataaataatgttaa 1397
 1524 aagctgaatgaagcagactagagtgatcatttaccctttaaataaataatgttaa 1583
 1398 ttgcagcgttgaagcagactagagtgatcatttaccctttaaataaataatgttaa 1457
 1584 ttgcagcgttgaagcagactagagtgatcatttaccctttaaataaataatgttaa 1643
 1458 tggatcctcaatcaacaaagaaagtgagtcagaaatgaagaaatgaagaaatgaagaa 1517
 1644 tggatcctcaatcaacaaagaaagtgagtcagaaatgaagaaatgaagaaatgaagaa 1703
 1518 ttccaataaatttcttcttcttcaataaataaataaataaataaataaataaataa 1563
 1704 ttccaataaatttcttcttcttcaataaataaataaataaataaataaataaataa 1749

RESULT 2
 US-08-277-540-2
 Sequence 2, Application US/08277540
 Patent No. 5474901
 GENERAL INFORMATION:
 APPLICANT: Drayna, Dennis T., Eaton, Dan L.
 TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:

```

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genetech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-277-540-2

Query Match      87.2% Score 1372: DB 1: Length 1749:
Best Local Similarity 90.3% Pred No. 0:
Matches 1558: Conservative 0: Mismatches 5: Indels 163: Gaps 2:

OY 1 aagaatgctgttggaagccttgcagccttgacgtccttgcacatgtctct 60
DB 24 AAAAAATGCTGTGGAGTGAAGCTTGGACGCTTGACGCTTGTACCATGTCTCT 83
OY 61 tctgtgagcagatgtcttcgcgttcacagatgagcgaattcttagctctctcagaa 120
DB 84 TCTGTAGAGAGATGCTTCGCGTTTCAAGTGGCCAACTTCACTGCTTCCAGAA 143
OY 121 cctctaggaagcttcaagcttctacagatcttctacacatatgatgtctctgc 180
DB 144 CCTCTAGGCAAGTTCAAGTTCTACAGAACTTACTACACATATGAGATTGCTCTGGC 203
OY 181 agcggtaacagctgaacctatctgtaagaaacaaagtcacatttctgtaaatgat 240
DB 204 ACCCGGTACAGCTGACCTTATTTGTGAAGAAAAACAAGTCATTTTTTGTAAATGCAAT 263
OY 241 ctgagtgcacaatgtgaagcccatcttaagtgcaggaattccatgcagtgtcttc 300
DB 264 CTGATGTCAACATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTGC 323
OY 301 tggcgaagctggaagatctctatccaacagacagattccaacgacagtcagcccgag 360
DB 324 TGGCAGACCTGGAAGATCTTATTCAACACAGACATTTCCAAACGACAGTACGCCCGGAG 383
OY 361 cctccgcacgtactactgaacagatcactcaactaaatgaatctatcttgatagaat 420
DB 384 CCTCCCATGCTACATGTGAAGATCACTCACTAAATGAATCTATCTTGTGATGAAT 443
OY 421 ttataactgagggcactctgatatgtcttaacaaatccaacatgtgatactctcattgaga 460
DB 444 TTATAACTGAGAGGCACTGTGATATGCTTACAAAATTCACACATTTGATCCTCATTTGAGA 503
OY 481 attacccactctatgtcttaaaagctcttgaaagaaacaaacagcgaataatgcatat 540
DB 504 AGTACCCACTCTATGTTTAAAGCTTTCGAAAAAAGAACACAGCAAAAATGCCATAT 563

```

```

OY 541 ggaatgactgtggaatccatgcagagaaatggaatctctcctgtctctgtctgtgtca 600
DB 564 GGATGACTGTGGAATCCATGACGAGAAATGATCTCTCTGCTTCTGCTGTGATCA 623
OY 601 taagccat----- 608
DB 624 TAGCCCATATTAACATCTGATGGAATAATAGGCAATATACCAATCTCTGAGGCTTG 683
OY 609 -----a 609
DB 684 TGGATTCTATGTTATGCGGCTGTTAATGTGACGCTTATGCTACTCATGAAAAAGA 743
OY 610 atcgaatgtggaagaaacgcttctctctatgcgaacaaatctgcacagaaagc 669
DB 744 ATGGAATGTGGAAGAAAGAACCGTTCTTTATGCGAAGCAATCATTTGATGGAAGAC 803
OY 670 tgaatagcaacttctctccaacaactggtgtgagaaggtgcacccagttcctcatgct 729
DB 804 TGAATAGGAACCTTGTGCTCAAAACACTGGTGTGAGGAAGGTGCATCCAGTTCTCATGCT 863
OY 730 cggaaacactactgtggaacttactcctgaagtcagaacagaaagtgaagcgagtgatgt 789
DB 864 CGAAACCTACTGTGACTTATCTGAGTCAAGACAGAAAGTGAAGCGAGTGCTAGTT 923
OY 790 tcttgaagaaatatacaacagatlaaagcatatcaagcatgcatctactccagc 849
DB 924 TCTTGAGAAAGAAATATCAACACAGATTAAAGCATACATCAGCATGATCTCCAGC 983
OY 850 atatagtttccatattctctatacagaaagtaagaagcaatgagaactgtctc 909
DB 984 ATATAGTGTTCATATCTCTATACAGAAAGTAAAGCAAGCAATGAGAACTGTCTCT 1043
OY 910 tagtagccagtgaaagagcttcgtctatctgacaaactagtaaaatataccaggtataac 969
DB 1044 TACTAGCACTGAAAGCACTGCTGCTATTTAGAAACATTAATAATACAGGTATACAC 1103
OY 970 atggcattggtctcaagaacctatacctagctcctcgtgaggtggaagatgagatcatg 1029
DB 1104 ATGGCATGCTGAGAAACCTTATACCTAGCTCTGAGGTGGAGATTTGATGATCATG 1163
OY 1030 atttggaatcaaatatcgt----- 1050
DB 1164 ATTTGGCATCAAAATATCTCTTACAAATGGAACCTTCAGATACGGGCACATACGATCT 1223
OY 1051 -----tiacataaacccacctgttagagaagctttgcccgtctctaa 1097
DB 1224 TGCTGCCGAGCGTTTACATCAAAACCCACTGTAGAGAACTTTTGCCGCTGCTTAA 1283
OY 1098 tagcttgcaatgcatatgaatgttaatgcccctgatttcaatctcgtctcgat 1157
DB 1284 TAGCTTGCAATGTCATATGGAATGTTTAATGCCCTGATTTTATCATCTGCTCCGAT 1243
OY 1158 ttaaatctactgattccagaaagcaaatcaatctgtatcaatatttttaagtttctc 1217
DB 1344 TTTAAATTTACTGATTCAGAGAACCAAAATCAATGTGATGATTAATTTAAATTTTATC 1403
OY 1218 cgtagtttgataaagatttctctatctgtgtctgcagagaacctaataaggtct 1277
DB 1404 CGTAGTTTATATAAGATTTTCTATCTTCTGTTCTGTGCAAGAACTTAATTAAGTCT 1463
OY 1278 acttgccattaaagcagactaggtatcagctcttttaccctttaaaaaaaatgttaa 1337
DB 1464 ACTTGCCATTTAAGCGAGCATAGGCTCATGTCGTTTATCCCTTTAAAAAAATTTGTA 1523
OY 1338 aagttcagtaactacttttcttgaatttgacagctttgactagcactatcaagcaact 1397
DB 1524 AAGTCTAGTACTACTCTTCTTTTGTGATTTTGAAGTTTACAGCTTATGCAAGCAACT 1583
OY 1398 ttgcagctttgactagcactatcaagcaagtttaataaagatcatctcaagctgcatat 1457
DB 1584 TTGACGCTTTGACTAGCATCTCAAGCAAGTTTATATCAAGATCATCTCATCGCTGATCAT 1643

```

QY 1458 tggatcctactcaacaaaggaaggtgicagagtaactaaagattctgtcccaa 1517
Db 1644 TGGATCCTACTCAACAAAGGAGGTGTCAGAGTACATTAAGATTCTCTCCAAA 1703
QY 1518 ttccaataaattctctctctcttcaaaaaa 1563
Db 1704 TTTTCATTAATTCTTCTCTCTTTAAAAA 1749

RESULT 3
US-08-430-787A-2
Sequence 2, Application US/08430787A
Patent No. 5593674

GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91

ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168

SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-430-787A-2

Query Match 87.2%; Score 1372; DB 1; Length 1749;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

QY 1 agaaaatgctgttggaagcttgcagccttgacgtctgttaaccattgttctct 60
Db 24 AGAAAATGCTGTGGAGTGAAGCTTGCAGCTTGACGTCTGTGACCATTTCTCT 83
QY 61 tctgtgagcagcagcttctgcgcttcagagtgagcgaagcttctgctgctctctaa 120
Db 84 TCTGTGAGCAGCAGTCTTCTGCGTTTCAGAGTGGCAAGTTCTAGCTCTTCTCGAA 143

QY 121 cctctgagcaagttcaagttctacagaatcttactacaacatactgagattgtctctgc 180
Db 144 CCTCTAGGCAAGTTCAGATTCTACAGAACTTACTACAAATATGAGATTGTCCTGCGC 203
QY 181 agccggttaacagcttgacctatctgtgaagaaaaaacagtcatttttjtaaatgcat 240
Db 204 AGCCGGTAACAGCTGACCTTATTTGTAAACAAAACAACTGATTTTGTAAATGCAAT 263
QY 241 ctgattgcaagcttgaaagcccaattaaatgtgagcggaaattccatgctgctctgc 300
Db 264 CTGATGTCGACATGTGAAGGCCATTTAAATGTGAGCGGAATTCAGTGTGCTTGC 323
QY 301 tggcagcgttggaagatcttcttaacagcagatttccagacacagtcagcccgag 360
Db 324 TGGCAGACCTGGAAGTCTTATTTCAACAGCATTTCCAAACACACAGTCACGCCCCGAG 383
QY 361 cctccgacatgctactatgacagatctcctcctaataatgaatctctcttgatagat 420
Db 384 CCTCCGATCGTACTATGAGAGTATCACTCACTCAATGAATCTATTTCTTGATGAAAT 443
QY 421 ttaatactgagagcagctctgatactgcttaacaaaaatccaatlgtgattccattgaga 480
Db 444 TTATACCTGAGAGCATCTCTGATATGCTTACAAAATCCAGATTGATCTCATTTGAGA 503
QY 481 agtaccacatctatgtttaaggttcttggaagaaacaaagcccaaaatgccaat 540
Db 504 ACTACCACTCATTTTAAAGTTTCTGCAAAAAGAACAAACAGCCAAAATGCGCATAT 563
QY 541 ggaattgactgtggaatccatgacagaaatgatatctctctctctgtctgtgttca 600
Db 564 GGATTGACTGTGGAATTCATGCGCAGAGATGATGATCTCTGCTTGTGTTGTTGTTCA 623
QY 601 taagccat----- 608
Db 624 TAGGCATATACATCAATTTCTATGGGAATTAAGGCAATATACAAATCTCTGAGCTTG 683
QY 609 -----a 609
Db 684 TGGATTCTATGATTATGCGGGTGTATATGAGAGGTATGACTACTCATGAAAAAGA 743
QY 610 atcgaatgtggaagaaagacgcttctctctctacgaaacatctgacatcggaacagacc 669
Db 744 ATCGAATGTGGAAGAAAGACGGTCTTCTATGGAACAAATCATGATGCGAAGACAGC 803
QY 670 tgaatgaacactgtctcctcaaacacatgtgtgtggaagtgatcgaatctctcagct 729
Db 804 TGAATGAGAACTTTGCTTCCAAACACATGCTGTGAGAAAGTGACATGCTTCATGCT 863
QY 730 cggaaacctactgtgacattatctcctgagtcagaaacagaaagtgaaagcagtgtagt 789
Db 864 CGAAACCTACTGTGACCTTATCTGATGACAGAACCAAGAAAGTGAAGCAGTGGCTAGTT 923
QY 790 tcttgagaagaataataaaccagtttaagcatatcaacagatgacatcttaactccagc 849
Db 924 TCTTGAGAAAGAAATATCAACAGATTAAGCATATACATCACATGATCTTACTCCAGC 983
QY 850 atatagtttccatctctctacacagaaagtaaaagcaagacatgaaagaaatctctc 909
Db 984 ATATAGTGTTCATATTTCTATACAGAACTAAACCAAGCAATGAGAACTGTCTC 1043
QY 910 taatgaagcagtgaaagcagctgtgctatgtgacaaactagtaaaatatacaggtatatac 969
Db 1044 TAGTAGCCATGGAAGCGTTCGTGTATGAGAACTAGTAAATAATACAGATATACAC 1103
QY 970 atgagcagctcagaagacattacatgactctctgaggttgagagcgtttgatactag 1029
Db 1104 ATGGCCATGCTCAGAAACCTTATACCTTACCTCTGAGGTGGAGAGTGTGATATAG 1163
QY 1030 atttgagcacaatattgt----- 1050
Db 1164 ATTGGGCAATCAATATTTGTTTACATTTGAATTTGAGATNCGGCAATACGAGATTCT 1223
QY 1051 -----tatacataaacccaactgttagagaagccttgcgctgtctctataaa 1097


```

Db 1224 TCGTCGCGAGGCTTACATCAAAACCCACCTGTAGAGAACCTTTGGCGTCTCTAATAA 1283
Oy 1098 tagcttgcatgtcattagaaatgttaatgccccgattatcatctgtctccgta 1157
Db 1284 TAGCTTGCGATGTCATAGGAATGTTAATGCCCTGATTTATCTTGTGTCGTAATC 1343
Oy 1158 tttaattactgattccagcaagaacaaatcgtatcatatattttaagttttatc 1217
Db 1344 TTTAATTTACTGATTCACACAGACCAAAATCATGTGATCAGATTAATTTTAAAGTTTATC 1403
Oy 1218 cgtatgtttgataaaagatttccattccctgtgtctgtctcagaagaactataatgtct 1277
Db 1404 CGTATTTTGTGATTAAGATTTTCCATTTCTGTGTTCTGACAGAACCTAATAAGTGTCT 1463
Oy 1278 acttgccatgaagcagacagtaggggttcacgtcttttaccctttaaanaaatgtta 1337
Db 1464 ACTTGTGCAATTAAAGCAGACGTAGGGTTCATGTCTTTTACCTTTAAAAAATTGTAA 1523
Oy 1338 aagctagttactactcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1397
Db 1524 AAGCTAGTACTACTACTTCTTCTTGTGATTTTGTGACCTTGTGACCTGACATCTCAAGCACT 1583
Oy 1398 ttgcagctgttgactgacatctcccaagcaagtttaatacaagaatcatcagctgtatcat 1457
Db 1584 TTCGACGTTTGTACTGACCATCTCAAGCAAGTTTAACAAGATCATCTACGCTGATCAT 1643
Oy 1458 tggactcctactcaaaaagaagagtggtcagaagttacatcaaatgatttctgtctcaaa 1517
Db 1644 TGGATCTTACTCAACAAAAGGAAGGGTGTCAAGAGTACATTTAAAGATTTCCTGCTCAAA 1703
Oy 1518 tttaataataattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1563
Db 1704 TTTTCAATTAATTTCTTCTTCTCTTAAATAAAAAA 1749

RESULT 4
US-08-869-057-1
; Sequence 1, Application US/08869057
; Patent No. 5985562
; GENERAL INFORMATION:
; APPLICANT: Morser, Michael J
; APPLICANT: Nagashima, Maiko
; TITLE OF INVENTION: Method of Detecting Thrombotic Disease
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Berlex Biosciences Legal Department
; STREET: 15049 San Pablo Avenue
; CITY: Richmond
; STATE: California
; COUNTRY: USA
; ZIP: 94804-0099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/869,057
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: A35
; ATTORNEY/AGENT INFORMATION:
; NAME: Washlien, Wendy L
; REGISTRATION NUMBER: 36,301
; REFERENCE/DOCKET NUMBER: 51509AUSM1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-262-3411
; TELEFAX: 510-262-7095
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1272 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PUBLICATION INFORMATION:
; AUTHORS: Eaton, Dan L.
; AUTHORS: Malloy, Beth E.
; AUTHORS: Tsai, Siao P
; AUTHORS: Henszel, William
; AUTHORS: Drayna, Dennis
; TITLE: Isolation, Molecular Cloning, and Partial
; TITLE: Characterization of a No. 5985562el Carboxypeptidase B
; JOURNAL: J. Biol. Chem.
; VOLUME: 266
; ISSUE: 32
; PAGES: 21833-21838
; DATE: No. 5985562 15-1991
; US-08-869-057-1

Query Match 58.4%; Score 918; DB 2; Length 1272;
Best Local Similarity 86.8%; Pred. No. 7 2e-213;
Matches 1104; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

Oy 18 atgaagctttgcagccttcgacgtccttgcacccattgtctcttctgtgagcagcatgtc 77
Db 1 ATGAAGCTTTGCAGCCTTGACGCTTGTACCATTTGTTCTTCTGTGACGACATGTC 60
Oy 78 ttgcggttccagagtggtccaaagtcttagctgtcttccagaaccccttgagcaagtctaa 137
Db 61 TTCGCGTTTCAAGAGTGCCCAAGTTCATGCTCTTCTTAAGACCTTAGCGAAGTTCAA 120
Oy 138 gtctcagaagcttactcaacaatagtattgtctctgtgcagccggtgaacagctgtac 197
Db 121 GTTCTACAGATCTTACTCAACATATGAGATTGTTCTCGGACGCCGATGAAGCTGAC 180
Oy 198 ctattgtgagaagaacaagaagtcacatlltttgttaaatgcatctgaltgcacaatlygt 257
Db 181 CTATTGTGAGAGAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGAATGTG 240
Oy 258 aaagccatttaattgtgaggaattccatgtaagtgctgtctgtgcagacgtggaagat 317
Db 241 AAAGCCATTTAATGTGAGGGAATTCATGAGTGTCTGTGCGACAGCGTGAAGAT 300
Oy 318 ctattcaacaagcagatttccacagcaacagtcagcccccagccctccgcatctactat 377
Db 301 CTATTCAACAGCAGATTTTCCACAGACAGCTAGCCCCCGACCTCCGATGTAAT 360
Oy 378 gaacaglatcactactaaatgaatctatcttctgtatagaaatttaactgagagcat 437
Db 361 GAACAGTATCCTCACTAATGAATCTATTTGTGATGAATTTAATCACTGAGAGGCAT 420
Oy 438 cctgatatgcttcaaaaatccacatgttgatctctcatlttgagaagtaaccactatgtt 497
Db 421 CCGATATGCTTACAAAACCCACATTTGATCTCATTTGAGAGTCCACCTATATGT 480
Oy 498 ttaaaagtctcgtaaaagaacaacaagccaanaaatgcatatggtatgttgatgtgaac 557
Db 481 TTTAAGGTTTCTGGAAGAAAGAACACAGCCAAAATGCCATATGATGACTGTGGAATC 540
Oy 558 catgcagagaatgagatctctctgtcttctgtctgtgtgtgltacaaagcat----- 608
Db 541 CATGCCAGAAATGATGATCTCTCTGCTTGTGTTGTGTTGATAGGCCATATATAACTCAA 600
Oy 609 ----- 608
Db 601 TTTATAGGATAATAGGGCAATATACCAATCTCTAGGCTTGTGATTTCTATGTTATG 660
Oy 609 -----aatcgaatgtggaag 626
Db 661 CCGGTGTTAATGTGACGCTTATGACTACTCATGAAAAAAGATGCAATGTGAGAGAAAG 720
Oy 627 aaccgttcttctatgcgaacaatcatltgcatcggaacagacctgaaatgacaattgttc 686

```

```

Db 721 AACCGTCTTATGCGAACAATCATTCGACGGAACAGACCTGATGGAACCTTTGCT 780
QY 687 tccaaacacgtggtggaagagtgacacagttccatgctcgaacacacgtgga 746
Db 781 TCCAAACACTGGTGTGAGAGAAAGTGCATTCATCTCTGCTCGAACCCTACTGTGGA 840
QY 747 ctttcctgagtcgaacagaagtgaagcagtggtgcttcttcgaagaataatc 806
Db 841 CTTTATCCGACTGAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTGAGAGAAATATC 900
QY 807 aaccagattaaagcattacatcgcatctatctctccagcatatagtgcttccatat 866
Db 901 AACCGATTAAGCATACATCAGCATTCATCTACTCCACCATATAGTGTTCATAT 960
QY 867 tccatacagaagaataagcaaacagcagtgagacgtgctcctgaagcagtgagca 926
Db 961 TCCCTATACAGCAAGTAAAGCAACCAATGAGAACTGTCTCTGATGACCCAGTAAACA 1020
QY 927 gtctgtctattgacaaactagtaaaataccaggtatatacatgycatgctcagaa 986
Db 1021 GTTCTGTCTATGAGAAAACCTGTAATAATACAGGTATACATGCGCATGCTCAGAA 1080
QY 987 accatacactagctcctcggagggtggagagatgagatcattgagtgatcgaataat 1046
Db 1081 ACCTTATACCTAGCTCTCTGAGAGGTGGAGCATGATGATGATTTGGCATCAAAAT 1140
QY 1047 tctgt-----ttac 1054
Db 1141 TCGTTTACATTTGAACTTCGAGATACGGGCAATACGATTCCTGCGGAGCGCTTAC 1200
QY 1055 atcaaacaccacgtgtaagaagctttgcgctgtctcttaataatagcttgacatc 1114
Db 1201 ATCAAAACCCACTGTAGAGAGCTTTTGCCTGTCTTAAATAGCTTGGCATGTCAAT 1260
QY 1115 aggaatgttaa 1126
Db 1261 AGGATGTTTAA 1272

```

RESULT 5
 US-08-782-760-5
 : Sequence 5, Application US/08782760
 : Patent No. 5948668
 :
 : GENERAL INFORMATION:
 : APPLICANT: Hartman, Jacob
 : APPLICANT: Fulga, Netla
 : APPLICANT: Mendelovitch, Simona
 : APPLICANT: Gorecki, Marian
 : TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
 : NUMBER OF SEQUENCES: 8
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Cooper & Dunham LLP
 : STREET: 1185 Avenue of the Americas
 : CITY: New York
 : STATE: New York
 : COUNTRY: U.S.A.
 : ZIP: 10036
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/782,760
 : FILING DATE: 13-JAN-1997
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/378,233
 : FILING DATE: 25-JAN-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: White, John P.

```

: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 0336/43847
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 927 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..927
: US-08-782-760-5

```

Query Match 7.9%; Score 124; DB 2; Length 927;
 Best Local Similarity 57.9%; Pred. No. 3.6e-21;
 Matches 278; Conservative 0; Mismatches 195; Indels 7; Gaps 3;

```

QY 613 gaatgtagaagaagaccgtctcttctctatgcaacaatcatgcatcggaacagactga 672
Db 362 GAATGTGAGAGAAAACCGGCTCTACTATGCTGGAGTTCCTGCTTGAGGTATGACCCCA 421
QY 673 atagaacttggtctcacaacactggtgtaggaagtgatccaagttccatcctcagg 732
Db 422 ACAGGAATTTTAATGC--TGGCTGTGTGAAGTGGAGCTTCGAGTCCCTGCTCTG 478
QY 733 aaactactgtagaacttatctctgagtcagacagagaagtaggaagcagtgtagttct 792
Db 479 AAACCTTACTGTGAGAACCCCAAGAGTCTGAAAAAGACAAAGCCCTGTGCAATTTCA 538
QY 793 tgagaagaataataacacagataaagcatatcagatgcatctactactccagcata 852
Db 539 TCCGCACMACCTCTCCACATCAAGGCTACTGACCATCTCACTACTACACAGATGA 598
QY 853 taggttccatctcctatcacagaagtaaaagcaagacatgaagaatgtctctag 912
Db 599 TGCTTACCTTACTCTATGACTACAAACCTGCTGAGAACTATGAGAAATTCATCC 658
QY 913 tagcagtgaaagcagttcgtctatgtgacaacactagtaaaataacaggtatacatg 972
Db 659 TGTGGAAGAGTGGCGCAAGAGCTTGCC--ACTTGCAATGCGACCAAGTACATATG 715
QY 973 gcatgtctcagaacacttaactagctcctggaagtgaggacgattgatactagat 1032
Db 716 GCCCAGGAGCTACAAACATCTATCTGCTGCTGGGAGATCGACGATGCTTATGATC 775
QY 1033 tgggcatcaaatatctgttaccatcaaacacacactgtaggaagctttgcgtgtc 1092
Db 776 AGGGAATCAATATTCCTTTACTTTGA-ACCTCGGAGTACAGGCTTCCTTGGCTTCTC 834

```

RESULT 6
 PCT-US96-00995-5
 : Sequence 5, Application PC/TUS9600995
 :
 : GENERAL INFORMATION:
 : APPLICANT: Bio-Technology General Corp.
 : TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
 : TITLE OF INVENTION: CARBOXYPEPTIDASE B
 : NUMBER OF SEQUENCES: 8
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Cooper & Dunham LLP
 : STREET: 1185 Avenue of the Americas
 : CITY: New York
 : STATE: New York
 : COUNTRY: U.S.A.
 : ZIP: 10036
 :
 : COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/00995
FILING DATE: 25-JAN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0336/43847-A-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 927 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..927
PCT-US96-00995-5

Query Match      7.9%; Score 124; DB 5; Length 927;
Best Local Similarity 57.9%; Pred. No. 3,6e-21;
Matches 278; Conservative 0; Mismatches 195; Indels 7; Gaps 3;

QY 613 gaatgtagaagaacgcttcttctatgcgaacatcatctgacggaacagctga 672
    |||||
DB 362 GAATGAGGAGAAAACCCCTCTACTATGCTGGAAGTTCCTGCTGGGTGAGACCCA 421
QY 673 atagcaacttgcttccaaacactggtgtagaagaagtgatccatgctcctcag 732
    |||||
DB 422 ACAGGAATTTTATGC---TGCGTGGTGGAAGTGAGAGCTTCTCGAGACCTGCTCG 478
QY 733 aaactactgtagactttatcctgtagtcagaacagaagtgaaagcagtgtagttct 792
    |||||
DB 479 AAACCTACTGTCGACCGACGCCAGAGCTCGAAGAAAGAGCAAGAGCCCTGCGAGATTGCA 538
QY 793 tgaagaagaatatcaacacagattaaagcatatcagcatgcatctatctccacagata 852
    |||||
DB 539 TCCGCACACACCTCTCCACCATTCAGGCTTACCTGACCACTCATCTACAGATGA 598
QY 853 tagtttccatattcctatacagaagtaaaagaacacagtaggaactgtcctcag 912
    |||||
DB 599 TCCTGTACCTTACTACTGATGACTACAACTGGCTGAGAACTATGAGGATTCGATGCC 658
QY 913 tagcgaagtagaagcagtgctgtctattgcaaaaactagtaaaaatccagggtatacag 972
    |||||
DB 659 TGGTGAAGAGTCCGGCAAGAGAGCTTGCC---ACTCTGATGCGACCAAGTACACATATG 715
QY 973 gcatggtcagagaacacattactagctctcgtgagtgtaggagacattgatctatgatt 1032
    |||||
DB 716 GCCCAGGAGACTACACATCTATCTCTGCTGCTGGGAGATGACGACTGCTTATGATG 775
QY 1033 tgggacataatattcgtttacatacaaacacacagctgtagaagaagctttgacgctgtctc 1092
    |||||
DB 776 AGGGAATCAAAATATCTTACTTTGA-ACCTCCGGGATACAGGCTTCCTTGGCTTCTTC 834
```

```

APPLICANT: Greenen, David P.
APPLICANT: Hersberger, Charles L.
APPLICANT: Larson, Jeffrey L.
APPLICANT: Steiner, Jane L.
APPLICANT: Zhang, Haichao
TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
PANCREATIC CARBOXYPEPTIDASE B
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/696,139
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,258
FILING DATE: 16-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-8681
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..921
US-08-696-139-3

Query Match      7.8%; Score 123.4; DB 1; Length 921;
Best Local Similarity 56.3%; Pred. No. 5e-21;
Matches 273; Conservative 0; Mismatches 206; Indels 6; Gaps 2;

QY 572 gatctccctgcttctgctgtggttcataagccataatcgaaatgtagaagaacg 631
    |||||
DB 321 GCTCATATTGATGCGTACTACTTACACTTGACCAAGACCAAGATGGAGAAAGCCG 380
QY 632 ttctttatgtagaacaatcatgcatcggaacagactgaatagaaacttgctccaa 691
    |||||
DB 381 CTCCTACCAATGCTGGAACCTACTGCTGACATGTCACAGAGACCCCAAGAAATTTGATGCG 440
QY 692 aacactggtgtagaagagtgcatccagttcctcatgctcggaacactactgtgacttta 751
    |||||
DB 441 G---TGGTGCAACACTGGAGCCCTCTACAGACCCCTGGATGAGACTTACGTGAGACTGCG 497
QY 752 tcttgatgtagaacacagaagtgagagtggtgctagtctcttgagaagaataatcaacca 811
    |||||
DB 498 TGCAGAGCTGGAAGAAAGAGCAAGGCCCTGCGGTGATTTATACGCACAACTCTCTC 557
QY 812 gattaaagcatatcagcatgcatcattactccagcatatagtggttccatattccta 871
    |||||
DB 558 CATCAAGACTACTGACGATCCATCATCTATCTTCCATCAGAGATGATCTTACCTTATTCCTTA 617
QY 872 tacacgaagtataaagaacacatgaggaactgctctctagtgtccagtgaaagcagttcg 931
    |||||
```

Db 618 TGATTACAACCTCCCGACAGACATCTGAGTTGAATTAACCTGGCTAAGGCTGCCGTGA 677
Qy 932 tgcattatgcaaaactagtaaaataccaggtatcacatgycatgctcagaacct 991
Db 678 AGAAGCTTG---CTACACTGTATGCGACCAAGTACACATACGCGCCAGAGCTACACACAT 734
Qy 992 ataccagctcccgaggaggaggagagatgcatctatgatttgggcatcaaatatcgtt 1051
Db 735 CTATCTGCTGCTGGGGGCTGTGATGACTGGCTTATGACCAAGAAATATTCTCT 794
Qy 1052 tacat 1056
Db 795 CACCT 799
RESULT 8
US-08-696-139-1
; Sequence 1, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hersiberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Hailiao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; NUMBER OF INVENTIONS: 6
; CORRESPONDENCE ADDRESS: PANCREATIC CARBOXYPEPTIDASE B
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1215 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1215
; US-08-696-139-1

Query Match 7.8%; Score 123.4; DB 1; Length 1215;
Best Local Similarity 56.3%; Pred. No. 5.5e-21;
Matches 273; Conservative 0; Mismatches 206; Indels 6; Gaps 2;
Qy 572 gatctctcgtcttctgcttggtcatalagccataatcgaatgtggaagaagacgc 631

Db 615 GCTCAATATTGATGGCTATACATCTACACCTGAGACCAAGAACCAATGTGAGAAACACCG 674
Qy 632 tctcttctatgcaaaacatcatcgcaggaagacctgaaatagcaacttgcctccaa 691
Db 675 CTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTGATGCTGG 734
Qy 692 acacgtgctgaggaaggtgcatccagttccatgctcgcggaacctctgtggaactta 751
Db 735 G---TGTTCCACAACCTGGAGCTCTTACAGACCCCTGCCATGAGACTTACTGTGGATCTGC 791
Qy 752 tccgtgaatgaagaacgaagtgagcgagtgctagtttcttgagaagaatlatcaacca 811
Db 792 TCCAGAGCTGAAAAAGAGACCAAGCCCTGGCTGATTTTATACCAACCAACCTCTCTC 851
Qy 812 gattaaagcatcatcagcatgcatctatctatccagcatatagtgttccatctcta 871
Db 852 CATCAAGCATACCTGACGATCCACTCATCTACACAGATGATCTTACCTTATTCCTA 911
Qy 872 tacagaagtaaaagcaagaacatgaggaactgtctctagtagccagtggaagatcgc 931
Db 912 TGATTACAACCTCCCGAGAACAAATGCTGAGTTGAATACCTGGCTTAAAGCTGCCGTGA 971
Qy 932 tgcattgcaaaactagtaaaataccaggtatcacatgycatgctcagaacct 991
Db 972 AGAAGCTTG---CTACACTGTATGCGACCAAGTACACATACGCGCCAGAGCTACACACAT 1028
Qy 992 ataccagctcccgaggaggaggagagatgcatctatgatttgggcatcaaatcgtt 1051
Db 1029 CTATCTGCTGCTGGGGGCTGTGATGACTGGCTTATGACCAAGAAATATTCTCT 1088
Qy 1052 tacat 1056
Db 1089 CACCT 1093
RESULT 9
US-08-860-882A-67
; Sequence 67, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGLESE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAH, JOHN FREDERICK
; APPLICANT: MARSHAM, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653

Db 672 GCTCAATATTGATGGCTACATCTACACCTGACCAAGAGCCGATTTTGAGAAAGACTCG 731
Qy 632 ttcttctatgcaacaatcattgcagcagaacagctgaatagcaacttgcctcaa 691
Db 732 CTCACCCCATGTGATCTAGCTGCTGACAGACAGCCCAAGAAATTTGATGC--- 788
Qy 692 acactggtgtggaaggtgcatccagttccatcgtcgtggaactactatgttgacttta 751
Db 789 TGGTGGTGTGAATTTGGAGCCTCTCGAAGCCCTGTGATGAAGAACTTACTGTGAGCTTCG 848
Qy 752 tcttgaagcagcaaggtgaaggcagtggttcttcttgaagaataatacaacca 811
Db 849 CGCAGAGTCTGAAAAGGACACCAAGCCCTGGCTGATTTTCATCCGACAAACTCTCTTC 908
Qy 812 gattaaagcatcaccagcatgcatcactccagcatatagtggttccatattcccta 871
Db 909 CATCAAGGATATCTGACATCCACTCGTACTCCCAATGATGATCTACCTTACTCANA 968
Qy 872 tacacgaagtaaaagcaagcagcagtgatgctcctcagtagcagtggaagcagtgctg 931
Db 969 TCTTACAACTCGGTGACAAACATGCTGATGAATGCCCTGCTAAAGCTACTGTGAA 1028
Qy 932 tgcatttgaacaaactagtaaaataccaggtatcacatagccatgctcagaaccc 991
Db 1029 AGACTTGGC---TCACCTGACGCGACCAAGTACACATATGGCCCGGAGCTACACAT 1085
Qy 992 ataccagtcctcgtgaggtggtgagcagattggtatgatttggtgagcacaatattcgt 1051
Db 1086 CATCTCTGCTGCTGGGGGCTCTGAGGACTGGCTTATGACCAAGAAATGATATTCCTT 1145
Qy 1052 tacat 1056
Db 1146 CACCT 1150

RESULT 13
US-08-860-882A-74
; Sequence 74, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1059 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-860-882A-74

Query Match 6.6%; Score 104.2; DB 2; Length 1059;
Best Local Similarity 53.8%; Pred. No. 2,3e-16;
Matches 261; Conservative 0; Mismatches 218; Indels 6; Gaps 2;

Qy 572 gactctccgcttctgtctgtgtgttcataagccataatcgatgtggaagaagacg 631
Db 387 GCTCAATATTGATGGCTACTTCTACACTGTGACCAAGAGCCGATTTTGAGAAAGACTCG 446
Qy 632 ttcttctatgcaacaatcattgcagcagaacagctgaatagcaacttgcctcaa 691
Db 447 CTCACCCCATGTGATCTAGCTGCTGACAGACAGCCCAAGAAATTTGATGC--- 503
Qy 692 acactggtgtggaaggtgcatccagttccctcaatgctcgtggaactactgttgacttta 751
Db 504 TGGTGGTGTGAATTTGGAGCCCTCGAAGCCCTGTGATGAAGAACTTACTGTGAGCTTCG 563
Qy 752 tcttgaagcagcaaggtgaaggcagtggttcttcttgaagaataatacaacca 811
Db 564 CGCAGAGTCTGAAAAGGACACCAAGCCCTGGCTGATTCATCCGAACTCTCTTC 623
Qy 812 gattaaagcatcaccagcatgcatcactccagcatatagtggttccatattcccta 871
Db 624 CATCAAGGATATCTGACATCCACTCGTACTCCCAATGATGATATCTACCTTACTCANA 683
Qy 872 tacacgaagtaaaagcaagcagcagtgatgctcctcagtagtgccagtggaagcagtgctg 931
Db 684 TGGTTCACAACTCGGTGACAAACATGCTGATGAATGCCCTGCTAAAGCTACTGTGAA 743
Qy 932 tgcatttgaacaaactagtaaaataccaggtatcacatagccatgctcagaaccc 991
Db 744 AGACTTGGC---TCACCTGACGCGACCAAGTACACATATGGCCCGGAGCTACACAT 800
Qy 992 ataccagtcctcgtgaggtggtgagcagattggtatgatttggtgagcacaatattcgt 1051
Db 801 CATCTCTGCTGCTGGGGGCTCTAAAGACTGGCTTATGACCAAGAAATGATATTCCTT 860
Qy 1052 tacat 1056
Db 861 CACCT 865

RESULT 14
US-08-860-882A-77
; Sequence 77, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2002, 15:21:57 ; Search time 1764.15 Seconds

(without alignments)
12034.527 Million cell updates/sec

Title: US-09-980-881-1

Perfect score: 1573
Sequence: 1 agaaaattgctgttgatg.....aaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

RSF: *
1: em_estba:*
2: em_estchum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	656	41.7	1446	11 AK004045	AK004045 Mus muscu
2	647	41.1	889	9 AL536069	AL536069 AL536069
3	622.8	39.6	780	10 BG616950	BG616950 602615625
4	615.6	39.1	806	10 BG567264	BG567264 602589745
5	604.2	38.4	788	10 BG618629	BG618629 602646064
6	581.2	36.9	747	10 BG566051	BG566051 602582652
7	579.2	35.7	725	10 BG618239	BG618239 602645967
8	561.8	35.3	738	10 BG616456	BG616456 602614618
9	555.2	35.3	735	9 AV646979	AV646979 AV646979
10	532.2	33.8	750	10 BG618813	BG618813 602646186
11	523.8	33.3	691	10 BG568240	BG568240 602587149
12	510.4	32.4	623	9 AV694058	AV694058 AV694058
13	509.4	32.4	583	9 AV693037	AV693037 AV693037
14	509.4	32.4	602	9 AV692032	AV692032 AV692032
15	507.4	32.3	621	9 AV698425	AV698425 AV698425
16	501.2	31.9	601	9 AV658390	AV658390 AV658390
17	491.6	31.3	643	9 AV688323	AV688323 AV688323

C	18	468	29.8	589	9	AV719484	AV719484
C	19	466	29.6	558	9	AV662298	AV662298
C	20	465.4	29.6	611	10	N63924	N63924 za28C09.s1
C	21	465	29.6	603	9	AM950700	AM950700 EST362770
C	22	464	29.5	586	9	AV719308	AV719308 AV719308
C	23	461.8	29.4	688	9	AV651709	AV651709 AV651709
C	24	460.2	29.3	495	9	AI271641	AI271641 q119h11.x
C	25	460.2	29.3	585	9	AV720894	AV720894 AV720894
C	26	460.2	29.3	631	9	AA700377	AA700377 z173b12.s
C	27	457.8	29.1	537	10	N98450	N98450 za28C09.r1
C	28	450.4	28.6	462	9	AV681868	AV681868 AV681868
C	29	442	28.1	576	9	AV720000	AV720000 AV720000
C	30	441.6	28.1	623	9	AV658178	AV658178 AV658178
C	31	438.6	27.9	586	9	AV719997	AV719997 AV719997
C	32	427.4	27.2	455	10	W88434	W88434 zb69f04.r1
C	33	425.2	27.0	575	9	AL536068	AL536068 AL536068
C	34	422.4	26.9	593	10	W89198	W89198 zb69f04.s1
C	35	420.6	26.7	747	10	BG565100	BG565100 602583720
C	36	418.4	26.6	450	9	AI076434	AI076434 o217b02.x
C	37	417.4	26.5	958	9	AI322676	AI322676 mj24d06.y
C	38	417	26.5	853	10	BI144644	BI144644 602909332
C	39	414.4	26.3	852	10	BG569281	BG569281 602588573
C	40	409.2	26.0	1022	10	BI332659	BI332659 602984408
C	41	405.6	25.8	544	9	BE235838	BE235838 143448 MA
C	42	397.8	25.3	530	9	AV720567	AV720567 AV720567
C	43	396	25.2	580	9	AI182113	AI182113 uc73f10.y
C	44	396	25.2	589	9	AA968307	AA968307 uc71h07.y
C	45	394.4	25.1	661	9	AI529872	AI529872 u183d11.y

ALIGNMENTS

RESULT	1	AK004045	1446 bp	mRNA	linear	HMC 19-JUN-2002
LOCUS	AK004045	Mus musculus	18 days embryo	whole body	cDNA, RIKEN full-length	
DEFINITION	AK004045	enriched library, clone:1110032P04:carboxypeptidase B2 (plasma), full insert sequence.				
ACCESSION	AK004045	AK004045.1	GI:12835067			
VERSION	AK004045					
KEYWORDS	HTC; CAP	trapper.				
SOURCE	Mus musculus (strain:C57BL/6J)	18 days embryo	cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library			
ORGANISM	Mus musculus					
REFERENCE	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P., and Hayashizaki, Y.					
TITLE	High-efficiency full-length cDNA cloning					
JOURNAL	Meth. Enzymol.	303, 19-44 (1999)				
MEDLINE	99279253					
PUBMED	10349636					
REFERENCE	2	(sites)				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.					
TITLE	Normalization and subraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes					
JOURNAL	Genome Res.	10 (10), 1617-1630 (2000)				
MEDLINE	2049374					
PUBMED	11042159					
REFERENCE	3	(sites)				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hatada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, Y., Yonekura, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuyama, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.					
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer					
JOURNAL	Genome Res.	10 (11), 1757-1771 (2000)				

Db 985 TACTGTTTCCATTCCTATACAGAAAGCAAGACAGACAGAACTGCTCTAG 1044
 QY 913 tagccagtgagagcagtcgtgctattgacaaactagtaaaataccaggtatacatg 972
 Db 1045 TGCCACGACGAGGAGTTCGTCGATTAAGCTATTAATAAACAACAGGATACACACG 1104
 QY 973 gcatagtcacgaacacttactagctctcgtgaggtgaggtgagatgtgatactatc 1032
 Db 1105 GCAAGTGCCTCAGAAAGCTTATATCTAGCTCTGAGGCTTCGACGATGTGATATGAT 1164
 QY 1033 tggagcacaataatctcgtt----- 1051
 Db 1165 TGGGCAATCAAAATTCGTTTACAATGAGCTCCGAGATACAGCAAGATTCCTGCG 1224
 QY 1052 -----tacatcaaacacacactgtagaagaagcttggcgcgtgctctaaatag 1100
 Db 1225 TGCCATGAGATATCAAAACCCACTGTCAGAAAGCTTGGCCGCGCATCTCTAAATAG 1284
 QY 1101 ctggagatgcatatgaatgtaatgccccgatttatacattcgtctcgatattc 1160
 Db 1285 TTTGGCATGTCATCAGAAACACTTAATGCCC-----TAACCTCCGCTCATTTATTTT 1338
 QY 1161 aattactgattccagcaagaacacatcatgtatcagat--tattttaagtattatcc 1218
 Db 1339 ATTTATGATTTTCAGCAACACTTAACCTGTGCAATAGCTTCAAGTTGAATCAGTTTC 1398
 QY 1219 gtagtttgataaagat 1236
 Db 1399 TTGCTTTTGTGAAGAT 1416

 RESULT 2
 AL536069 889 bp mRNA linear EST 13-FEB-2001
 LOCUS AL536069 LRI_F1013_Fbrn1 Homo sapiens cDNA clone CS0DF022YE21 5
 DEFINITION prime, mRNA sequence.
 ACCESSION AL536069
 VERSION AL536069.1 GI:12799562
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 889)
 Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..889
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DF022YE21"
 /clone_lib="LRI_F1013_Fbrn1"
 /dev_stage="Pooled tissue from post conception fetuses (20
 week, 24 week and 26 week)"
 /lab_host="DH10B"
 /note="Organ: Petal brain; Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(dt) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and cloned into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by Life
 Technologies. Contact : Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville
 , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
 fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"
 BASE COUNT 255 a 186 c 192 g 255 t 1 others
 ORIGIN

Query Match 41.1%; Score 647; DB 9; Length 889;
 Best Local Similarity 86.7%; Pred. No. 1,3e-85;
 Matches 771; Conservative 1; Mismatches 6; Indels 111; Gaps 1;

 QY 28 gcagccttgagtccttgtaaccattgtctctctctgtgagcaacagctctcgcttc 87
 Db 1 GCAGCCTTGAGTCCTTGATCCCATTTGCTTCTGTGACACAGATGCTTCCGCTTC 60
 QY 88 agaagtgcaagttcttagctgctctctcctagaaacctcaagcaagttcaattacaga 147
 Db 61 AGAGTGCCAAAGTTCAGCTGCTCTTCCCTAGAACCTTAGGCAAGTTCAGTTACAGA 120
 QY 148 atcttactacaacatagagattgtctctgagcccggttaacagctgacattatgta 207
 Db 121 ATCTTACTACAACATATAGAGATGTTCTGTCGCGACCGGTAAACAGCTGACCTTATGTGA 180
 QY 208 agaaaaaaccaagtcatttttttgaatgatcatgtatgctgcaaaagtgaagccact 267
 Db 181 AGAAAAAACAAAGTCCATTTTGTAAATGCATGTGATGACAAATGTGAAGCCCAT 240
 QY 268 taatgtgagcagatccatgcatgctgtctgagcaagctggaagatcttacaac 327
 Db 241 TAAATGAGCGGAATTCATGCGAGTGTCTTGCGGACAGAGTGAAGATCTTATTCAC 300
 QY 328 agcagattcccaagcacagtcagccccgagccctcgacatgctactatgaacagatc 387
 Db 301 AGCAGATTTCCAAAGCAGACAGTACGCCGCCGAGCTCGCATCTGATATGAACAGTATC 360
 QY 388 actactcaaatgaaatctatctctgtgatagaattataactgagaagcatcgtatagc 447
 Db 361 ACTCACTAAATGAAATATATCTTGGATGATGAAATTAATACGAGACACCTCTATATGC 420
 QY 448 ttacaaaatcacatctgagatcctcatttgagaagtaaccactatgitttaagatt 507
 Db 421 TTACAAAATCCACATTTGATGATCCATTTGGAAGTACCCACTATATGTTTAAAGTTT 480
 QY 508 ctggaagaagaacaaacagccaaatgcatatgattgactgtgaaatccatgccagag 567
 Db 481 CTGGAAGAAAGAACAGCAGCAGCAAAATGCAATGATGATGATGATGATGATGATGATG 540
 QY 568 aatgactctctcgtct 608
 Db 541 AATGATCTCTCCGCTTCTGCTTGTGTGTCATAGGCCATATACATCAATTTATGGA 600
 QY 609 ----- 608
 Db 601 TAATAGGCAATATACCAATCTCTGAGGCTTGTGATTTCTATGTTATGCCGTGTTA 660
 QY 609 -----aatgaatgtagaagaagaccgttctt 636
 Db 661 ATGTGATGTTATGACTACTCATGAGAAAAAGATGCAATGAGAGAGAAACGTTCTT 720
 QY 637 tctatggaacaatcatatgcatggaagacacccatgaatgaacttcttccaaacact 696
 Db 721 TCTATGGAACAATCATATGATGATGGAACAGCCATGATGGAACCTTCTCTCAAAACAT 780
 QY 697 ggttgaggaagtgatcagcttcctcctcatgctcgtggaacactactgtggaattatct 756
 Db 781 GGTGTGAGGAAGTGCTATCCAGTTCTCTCATGCTCTGGAACCTTACTGTGGACTTTATCTG 840
 QY 757 agtcagaacccaagtagagcagtgtagttctcttgaagaagaatat 805
 Db 841 AGTCAGAACCAAGAGTGAAGCAGTGTGCTGATCTTCTGAGAAAGAAAT 889

 RESULT 3
 BG616950 780 bp mRNA linear EST 18-APR-2001
 LOCUS BG616950 602615625F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4734247 5'
 DEFINITION mRNA sequence.
 ACCESSION BG616950
 VERSION BG616950.1 GI:13668321

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM1597 row: b column: 08
High quality sequence stop: 754.

FEATURES
Source
Location/Qualifiers
1..780
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4734247"
/lab_host="NIH_MGC_76"
/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgccgctggc); Site_2: SfiI (ggcgccatagcgc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGCGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGAGCGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 237 a 147 c 146 g 250 t
ORIGIN

Query Match 39.6%; Score 622.8; DB 10; Length 780;
Best Local Similarity 91.5%; Pred. No. 4,6e-82;
Matches 703; Conservative 0; Mismatches 12; Indels 53; Gaps 2;

QY 850 ataatgtttccatattccctacacgaagaagaacacatggaactgtctc 909
Db 1 AATAGTGTTCATATTCCTATACAGAAATTAAGTAAATACAGATATAC 60
QY 910 tagtagcagtagaagcaattcgtctattgacaaactagtaaaataccagataaac 969
Db 61 TAGTAGCCAGTAGACAGTTCGTGCTATTGAGAAAATTAAGTAAATACAGATATAC 120
QY 970 atggcctagctcagaaacctatatacctagctcctggaggtggagagatgcatatg 1029
Db 121 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCGAGAGTGGGACATTTGATCTATG 180
QY 1030 atttgagcacaataatcgt----- 1050
Db 181 AATTGGGATCAAAATATTCGTATTACAACTTCGAGATACGGGACATACGATTCCT 240
QY 1051 -----ttacatcaaacaccctctgtagaagaagctttgcgcgtctctaa 1097
Db 241 TGTCTCCGAGCGGCTTACATCAAAACCTCTGAGGAAGCTTTGCGCTCTCTTAA 300
QY 1098 tagcttgcatgcatlagaatgtltaatgcccctgattatcatctgcttcgtat 1157
Db 301 TAGCTTGCGATGATATGGAATGTTTATGCCCCCTGATTTATCATCTGCTTCGAT 360
QY 1158 tttaattacgattcccgacgaagacaataatgatacgaatatcttaagtattac 1217
Db 361 TTTAATTACTGATTCCAGCAAGACCAATATGATATTTTAAAGTTTATTC 420

QY 1218 cgtagtttgataaagaatttccatattccttggtctgctcagagaacataaagtct 1277
Db 421 CGTAGTTTGAATAAGATTTTCCATTCCTGTTCTCTGACAGAGAACTAATTAATGCT 480
QY 1278 acttgccattaaagagactaggtgtcattcttttccctttaaataaattgttaa 1337
Db 481 ACTTGCCTAATAAGGACAGACTAGGTTCAATGCTTTTACCCCTT-AAAAAAATTGTAA 539
QY 1338 aagtcagttacactttcttcttgatttcagcttgtagtaagcacttaagaact 1397
Db 540 AAGTCTAGTACTACTTTTCTTTGATTTTGACACTTTGACATCTCAAGCAAGT 599
QY 1398 ttgcagctttgactagcactcctcaagaagtttaalcaagaalcatctcagctgcat 1457
Db 600 TTGCAAGTTTGAAGCTACATTCACAGCAAGTTTAATCAATGATTCACGCTGATCAT 659
QY 1458 tggatctctctcaacaaaggaaggtgtgtcgaagtaacataaagattctgtccaaa 1517
Db 660 TGGATCTTACTCAACAAAGGAGGTTGTCAGAGTACATTAAAGATTTCGTCCAAA 719
QY 1518 ttctcataattcttcttctctcctttaaataaataaataaataa 1565
Db 720 TTTTCAATTAATTTCTGTTGCTTACTTAAATTAATTAATTAATTAATTAATTAAT 767

RESULT 4
BG567264
LOCUS 806 bp mRNA linear EST 10-APR-2001
DEFINITION 602589745r1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4723690 5',
mRNA sequence.
ACCESSION BG567264 GI:13574917
VERSION BG567264.1 GI:13574917
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM1583 row: j column: 11
High quality sequence stop: 772.

FEATURES
Source
Location/Qualifiers
1..806
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4723690"
/lab_host="NIH_MGC_76"
/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgccgctggc); Site_2: SfiI (ggcgccatagcgc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGCGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGAGCGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 239 a 158 c 162 g 247 t
ORIGIN

Query Match 39.1%; Score 615.6; DB 10; Length 806;
Best Local Similarity 90.7%; Pred. No. 5,1e-81;
Matches 724; Conservative 0; Mismatches 14; Indels 60; Gaps 4;

614 aatgtaggaagaagaaccgtcttcttctatgcgaacatcatctatcatgcgaagactgaa 673
1 AATGTGGAGAAAGAACCGTTCTTTCTATGCGAACATTCATGCAATCGGAACAGACTGAA 60
674 taggaactgtctccaaacactgggtgtaggaaggtgcatcgaatctctcctcga 733
61 TAGGAACCTTGTCTCCAAACACTGGGTGAGGAAGGTGATCCAGTCTCTACGTCGGA 120
734 aaactactgtgacttattctctgagtcagaaacgaagtgaaaggaagtgacttctt 793
121 AACCTACTGTGACTTATCTGAGTCAAGAACAGAAAGTGAAGGCGCTAGTCTTCT 180
794 gagaagaataatcaaccagaattaaagcatatcaatcagatcagatcttaccagatat 853
181 GAGAAAGAAATATCAACAGATTAAGCATATCAGCATGATTCATCTCCAGCATAT 240
854 agtcttccatattcctatatacagaaglaaagaacacatagagaactgtctctagt 913
241 AGTGTTCATATTCCTATACAGAGAAAGTAAAGCAAGCAATGAGAACTGTCTAGT 300
914 agcagatgaagcagatctgtgtctatgacaacaaactaglaaataaccaggtatacactg 973
301 ACCCGATGAGAGAGTTCGTGCTATGAGAAACTAGTAAATACCCAGGTATACACATG 360
974 ccatgagcagaagaacctatctatcagctcctgaggtgaggtgagcagatgactatgatt 1033
361 CCAATGCTGAGAACTTATACCTAGCTCTGAGAGGTGGAGAGATGATATATATTT 420
1034 gggcatcaaatatcgt----- 1050
421 GGGCATCAAAATATGCTTTACATTCAGATTCGAGATAGGGACATAGAGATTTCTGCT 480
1051 -----ttacatcaaaacccactgttagagaagctttgcgcgtgtctctaaatagc 1101
481 GCCGGAGCCTTATCATCAAAACCCAGTGTAGAGAAAGTTCGCGCTCTCTAAATATAG 540
1102 ttggcagatcatatgaagaatgttaatgagccctgatttcatcttgcctcgtatttta 1161
541 TTGGCATGTCATTAGGAATGTTTAATGCCCTGATTTTATCATTCGCTTCCGTATTTTA 600
1162 attactgattccagcaagaacaaatcattgatatcagattattttaagtttatccgta 1221
601 ATTACTGATTCAGAGAACCAAAATCATGATGATTTATTTTAAAGTTTATCCGTA 660
1222 -gtttgtataaaagatttccattccctgtgtctgtctgtagagaacactaataagtcact 1280
661 GGTATTGATMAAGATTTTCCATTCCTGTTCTGTCTGAGAGAACCTAATAGTGTAA-- 718
1281 ttggcacttaagcagaactaggggttcatgctcttaccctttaaataaataatgtaaaag 1340
719 -TTGCATTAAAGCAGACTAGGGTTCATGCTTTTATCCCTTTAAAAAAAATGT---AAAG 773
1341 tctagttactactttt 1358
774 TCTAGTTACTAATTTCTT 791

RESULT 5
Bg618629 788 bp mRNA linear EST 18-APR-2001
LOCUS Bg618629 602646064F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',
DEFINITION mRNA sequence.
ACCESSION Bg618629
VERSION Bg618629.1 GI:13670000
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 788)
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LNCMI630 row: 0 column: 15
High quality sequence stop: 760.
Location/Qualifiers

FEATURES
source 1.788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4767974"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDMR-LIB (Clontech); Site: 1: SfilI (ggccgcctggcc); Site: 2: SfilI (ggcattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTAATGAGCC-3' and 3' adaptor sequence: 5'-ATTCTGAGCGGAGCGGCGGCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC library."

BASE COUNT 222 a 157 c 169 g 240 t
ORIGIN

Query Match 38.4%; Score 604.2; DB 10; Length 788;
Best Local Similarity 99.5%; Pred. No. 2.4e-79;
Matches 606; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 agaaaattgctgttggaagaagctttgagccttgacgctctgttaaccattgttctc 60
49 AGAAATGCTGTGGAGAAAGCTTTCAGCCCTTCAGCTCTGTACCCATTGTTCTCT 108
61 tctgtagcagcatgtcttcgcttccagagtgagcgaagttcagctgtctctcctaga 120
109 TCTGTGAGCAGCATGCTTCTGCGTTTCAAGTGTGCGCAAGTTCAGCTCTCTAGAA 168
121 cctctagcgaagttcagaagttctcaagaattctactcaacatatgagattgtctctg 180
169 CCTCTAGGCMAAGTTCAAGTCTACAGAAATCTTACTACACATATGAGATGTTCTCTG 228
181 agccggttaacagctgaccttatgtgtgaagaaaaaagaatccatttttggtaagcat 240
229 AGCCGTTAACAGCTGACCTTATTTGTGAAGAAAAAAGCCATTTTGTGAAGCAT 288
241 ctgattcgacaattgtgaagccatttaaatgtgagcgggaattccatgagtgcttcg 300
289 CTGATGTGACATGTGAAGAAAGCCATTAAATGTGAGCGGAATTCATGCAAGTGTCTG 348
301 tggcagaagcttgaagaattctattcaacagcagaattcccaagcagacagtcagcccgag 360
349 TGGCAGATGTGGAAGATCTTATTCACAGCAGAGATTTCCAAAGCAGACAGTCAGCCCGAG 408
361 cctccgcatcgtactatgaaacagatcatcactcaataaagaaatcttcttgatagat 420
409 CCTCCGATGCTACTATGAAGACAGTATCCTCACTAATGAAATCTTCTTGGATGAAT 468
421 ttaatactgagagcactcctgatatgcttaaaaaatccacatctgagtcctcattggag 480
469 TTATTAAGTGAAGCATTCGATATGCTTACAAAATTCACATATGGATCCTCATTTGAGA 528
481 agtaccacactatgttttaaaagtttctgtgaaagaagaacacagcacaataatgcatat 540

```
|||||
Db 529 AGRACCACTCTATGTTTAAAGTTTCTGAAAAGAAAGAGAGCCAAAATGCCATAT 588
QY 541 ggaatgacgttggaatccatgacagagaatgatactctctgctcttctgtgttca 600
Db 589 GGATTCAGCTGTGAATCAGCAGAGAAATGATCTCTCTGCTTCTGCTTGTGTTCA 648
QY 601 taagccata 609
Db 649 TAGGCCATA 657

RESULT 6
Bg566051 747 bp mRNA linear EST 10-APR-2001
LOCUS 602382652F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710564 5',
DEFINITION mRNA sequence.
ACCESSION Bg566051
VERSION Bg566051.1 GI:13573704
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 747)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: L1CM1549 row: g column: 13
High quality sequence stop: 731.
FEATURES
source location/Qualifiers
1..747
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4710564"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site:1:
SfiI (ggcgccctcgcc); Site_2: SfiI (ggccatattgccc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATATGAGCC-3' and 3' adaptor sequence:
5'-ATTCAGAGCCGAGCGGCGGACATG-dr(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 212 a 157 c 159 g 219 t
ORIGIN
Query Match 36.9%; Score 581.2; DB 10; Length 747;
Best Local Similarity 99.2%; Pred. No. 5.6e-76;
Matches 605; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
```

```
|||||
Db 123 CCTATGGCAAGTTCAGATTTACACAACTTACTCAACATGATGATTGCTGTG 182
QY 181 agccggttaacagatcgaacctattgtgaagaaaaaagaagtcacattttgttaatgcat 240
Db 183 AGCGGGTAACAGGTGACCTTATGTGAGAAAAAACAAAGTCATTTTGTGAAATGCAT 242
QY 241 ctgattcgacacatgtgaagaccatttaaatgtgagggaaattccatgacgtctctgc 300
Db 243 CTGATGTCCACATGTGAAAGCCCATTTAAATGTGAGCGAAWTCATGACATGCTTGC 302
QY 301 tggcaagcttggaagatcttattcaacagcagatcttcaacagacacagtcaccccgag 360
Db 303 TGGCAGACGTGGAAGATCTTATTCAACAGCAATTTCCACAGCAGACATGACCCCGAG 362
QY 361 cctcgcatcgactcatgaacagatcactcaactaataatgaattatcttggatagaat 420
Db 363 CCTCGCATCGTACTATGAAAGATGACATCACTAAATGAATCTATTCTGGATAGAAAT 422
QY 421 ttataactgagagcattccctgatatgcttacaataatccattggatccatttgaaga 480
Db 423 TTATAAGTGAAGGATCTCTGATATGCTTACAAAATCCACATTTGATCTCATTTGAGA 482
QY 481 agtaccactctatgttt-taaagttcttggaagaaacaacagccaaaatgacata 539
Db 483 ACTACCACTCTATGTTTCTAAAGGTTCTG6AAAAGAAAGAGAGCCAAAATGCCATA 542
QY 540 tggattgacgttggaatccatgacagagaatgatactctctgctcttctgtgttc 599
Db 543 TGATTCAGCTGTGAATCAGTCAGAGAAATGATGATCTCTGCTTCTGCTGTGTC 602
QY 600 ataagccata 609
Db 603 ATAGGCCATA 612

RESULT 7
Bg618239 725 bp mRNA linear EST 18-APR-2001
LOCUS 602645967F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767376 5',
DEFINITION mRNA sequence.
ACCESSION Bg618239
VERSION Bg618239.1 GI:13669610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 725)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: L1CM1629 row: f column: 17
High quality sequence stop: 723.
FEATURES
source location/Qualifiers
1..725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4767376"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site:1:
SfiI (ggcgccctcgcc); Site_2: SfiI (ggccatattgccc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
```

sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGACGGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 203 a 153 c 149 g 220 t
ORIGIN

Query Match 36.8%; Score 579.2; DB 10; Length 725;
Best Local Similarity 99.2%; Pred. No. 1.1e-75;
Matches 603; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 agaaattgtctgtggaatgaagcttgcagccttcagctcctgtacccattgtctc 60
DB 48 AGAAATGCTGTGGATAGCTTGACGCTTGACGCTTGACGCTTGACGCTTGCTCT 107
QY 61 tctgtgagcagcaltcttcgcgttcagagtgagcgaagttcagctgtcttcctaga 120
DB 108 TCTGTGAGCAGCATGCTCTGCGCTTGACAGTGCGCAAGTTCTAGCTGCTCTCTGAA 167
QY 121 cctctgagcagatcctcaagcttcaagatcttcaacataatgaattgtctctggc 180
DB 168 CCTCTAGGCAAGTTCACAGATCTTACACATATGAGA-TGTTCTCTGGC 226
QY 181 agccggaagcagctgagccttctgtgaagaaagaaagctcattcttctgttaagtc 240
DB 227 AGCCGGAACAGCTGACCTTA-TGTGAAGAAAAACAAGTCATTCTTTGTAAATGAT 285
QY 241 ctgagtgcgaacatlygaagccattaaatlygaagcgaattcatalgcagtgcttc 300
DB 286 CTGATGTCGACAAATGTGAAGCCCATTTAAATGTGACGCGAATTCATGCAAGTCTTGC 345
QY 301 tggcagacgfggaagatcttattcaacagagatttccaaagaaacagtaagcccgag 360
DB 346 TGGAGATGTGGAAGATCTTATTCACAGAGATTTCCAAAGACAGTCAGCCCGGAG 405
QY 361 cctccgagctgactaagacagatcaactcaactaaatgaatcattctctgtgtaga 420
DB 406 CTCGCGATGCTACTAGAACAGATACACACATTAATGAATCTATCTTGATAGAT 465
QY 421 ttatactgagagcagctcgtatagcttcaacaaatccacattgagctcctcattaga 480
DB 466 TTATTAATGTGAGAGCATCTGATGCTTCAAAAATCCCATTTGATGCTCATTTAGA 525
QY 481 agtaaccacactatgttttaaggttctcgaagaaagaaagcgaagaaatgcatat 540
DB 526 AGTACCCACTATGTTTAAAGTTCTGGAAGAAAGCAAGCAAAAATGCCATAT 585
QY 541 ggaattgagctggaatccatgcagcagagatgagctcctcgtcttctgtgtgttca 600
DB 586 GGAATGACTGTGGAATCCATGCCAGAAATGAGATCTCTCCGCTTCTGCTTGATTCA 645
QY 601 taagccat 608
DB 646 TAGGCCAT 653

RESULT 8
LOCUS BG616456 738 bp mRNA linear EST 18-APR-2001
DEFINITION 602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',
ACCESSION BG616456
VERSION BG616456.1 GI:13667827
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 738)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@pds-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLML at:
<http://image.llnl.gov>
Plate: LNCM1595 row: C column: 05
High quality sequence stop: 666.

FEATURES
source
location/Qualifiers
1..738
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4733500"
/lab_host="NIH-MGC-76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1;
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGACGGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 209 a 161 c 153 g 215 t
ORIGIN

Query Match 35.7%; Score 561.8; DB 10; Length 738;
Best Local Similarity 99.6%; Pred. No. 3.9e-73;
Matches 563; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 45 gtaccattgttctctctctgtgagcagatcttcgctgcagagtgagcgaagttccta 104
DB 1 GTACCAATGCTTCTCTCTGTGAGCAGATGCTTCCGTTTCAGAGGCCAAGTCTA 60
QY 105 gctgccttctcctaagaccttaaggcaattcaagtlctacagaactctactacaacat 164
DB 61 GCTGCTCTCTCTAGACCTCTAGGCAAGTCAAGTCTCAAGATCTTACTACAACATAT 120
QY 165 gagattgttctctgagcagcgtaacagctgaacttatttgaagaaacaaagtcac 224
DB 121 GAGATTGTTCTCTGCGACCGGTAAAGCTTATTTGGAAGAAAAAAGTCCAT 180
QY 225 ttcttgaatgatcctgagctgagacagtgtgaagccatttaaatgtgagcgaatt 284
DB 181 TTTTGTGAATGATCTGATGATGTCACAAATGAAAGCCATTAAATGTAGCGGAAAT 240
QY 285 ccacgagtgctctgtcgtgagcagcgtggaagatcttattcaacagcagattccacgac 344
DB 241 CCATGCAATGCTTGTGCGAGAGCTGCAATGATGATGCAATGATGCAATGATGCAATG 300
QY 345 acaatcagccccgagcctcgcagctcgtactatgaacagatcactcaatgaatc 404
DB 301 ACAATGACCCCGAGCGCTCGCATGATGATGATGATGATGATGATGATGATGATG 360
QY 405 tattctgagatagattataacagagagagcagctcgtatgattacacaaatccacat 464
DB 361 TATCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 465 ggaatcctcatttgaagatccacactcattgttttaaggttctcgtgaagaaacaaaca 524
DB 421 GGATCTCATTTGGAAGATGACCACTCATGTTTAAAGGTTTCGGAAGAAAGCAAGCA 480
QY 525 gccaaaatgcatatgattgactgtggaatccatgcagagaaatgattctcccgct 584
|||||

REFERENCE 1 (bases 1 to 623)
 Mammalia: Eutheria; Primates; Catarrhini; Homiidae; Homo.
 AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
 Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
 Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
 Hu, G., Gu, J., Chen, Z., and Han, Z.
 TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 MEDLINE 21625106
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES
 source
 1. 623
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GKCADB09"
 /clone_1lb="GKC"
 /tissue_type="hepatocellular carcinoma"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 166 a 137 c 130 g 186 t 4 others
 ORIGIN

Query Match 32.4%; Score 510.4; DB 9; Length 623;
 Best Local Similarity 98.8%; Pred. No. 1.4e-65;
 Matches 514; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 gaaatctgctgttggaagcttgcagccttgagctctgtccattgtctctt 61
 Db 1 GAAATGCGTGGGATGAGAGCTTGACGCTTGACGCTTGACGCTTGCTTT 60
 QY 62 ctgtgagcagatgctctgcgttcagagtgagcagatcttagctctctctagaac 121
 Db 61 CTGTGAGCAGCATGCTTCGCTTCAGAGTGCCAAAGTCTAGTGCCTTCTTGAGAC 120
 QY 122 ctctaggcaagcttcaagcttaccagaatcttaccatattgagattgtctctgca 181
 Db 121 CTCTAGGCAAGCTTCAAGCTTCTACAGAACTCTACTACACATATGAGATTGTTCTTGCGCA 180
 QY 182 gccggttaacagctgacctatttgtaagaagaaacagtcacatttctgttaatgcatc 241
 Db 181 GCCGGTAAACAGCTGACCTATTGTGTGNAAMAAACAGTCCATTTTGTGTAATGCAATC 240
 QY 242 tgatgtcgacaatgtgaagccatttaagtgtgagcggaaatccatgcatgtctctgtct 301
 Db 241 TGATGTGCAAAATGTGAAGCCATTAAATGTGAGCCGAATTCATGACAGTGTCTTGCT 300
 QY 302 ggcagacgttgagaatcttaccacaagcagattccacaagcagacagtcagcccgagc 361
 Db 301 GGCAGACGTGGAAGATCTTATTCAACAGAGATTTCCAAACGACAGTCAGCCCGAGC 360
 QY 362 ctccgcatgtactatgaacagatcactactactaaatgaatctattctgatatgaatt 421
 Db 361 CTCCGCATCTGCTATGAAACAGTATCACTCAATGAATCTATTCTTGATGATGAAAT 420
 QY 422 tataactgagagcactctctgatatgtcttacaataaaatccacattgtgatttgagaa 481
 Db 421 TATAACTGAGAGCATCTCTGATATGCTTACAAAATTCACATTGATCTCATTTGAGAA 480
 QY 482 gtaccactctatgttttaaggttctcgaaagaacaa 521
 Db 481 GTACCACTCTATGTTTAAAGGTTTCTGGAGAACAGCA 520

RESULT 13
 AV693037
 LOCUS
 DEFINITION
 AV693037
 AV693037
 AV693037.1 GI:10294900
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 583)
 AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
 Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
 Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
 Hu, G., Gu, J., Chen, Z., and Han, Z.
 TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 MEDLINE 21625106
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES
 source
 1. 583
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GKCAFH12"
 /clone_1lb="GKC"
 /tissue_type="hepatocellular carcinoma"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 161 a 130 c 116 g 176 t
 ORIGIN

Query Match 32.4%; Score 509.4; DB 9; Length 583;
 Best Local Similarity 98.8%; Pred. No. 2e-65;
 Matches 513; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 aaatctgctgttggaatgaagcttgcagccttgagctctgtaccattgtctcttc 62
 Db 1 AAATGCTGTGGGATGAGAGCTTGACGCTTGACGCTTGATGCCATTTGTTCTTTC 60
 QY 63 tgtgagcagatgtcttcggttcagagtgagcagattcttagctctctctagaacc 122
 Db 61 TGTGAGCAGCATGCTTCTTGCTTCAGAGTGCCAAATTTCTAGCTGCTCTTCTTGAGAC 120
 QY 123 tctaggcaagcttcaagcttaccagaatcttaccatattgagattgtctctgagc 182
 Db 121 TCTAGGCAAGTTCAGATCTTACAGAACTTACAAACATATGAGATTGTTCTTGCGCA 180
 QY 183 ccggttaacagctgacctatttgtaagaagaaacaaagtcacatttctgttaatgcatc 242
 Db 181 CCGGTAAACAGCTGACCTATTGTGAGAAAACAAAGTCCATTTTGTGTAATGCAATCT 240
 QY 243 gatcgacaatgtgaaagccatttaaatgtgagcgaattccacagcagtgctctgtg 302
 Db 241 GATCGACAATGTGAAGCCCATTTAAATGTGAAGGGAATTCATGACAGTGTCTTGCTG 300
 QY 303 gcagagcttgagaatcttattcaacagcagattccacaagcagtcagcccgagc 362
 Db 301 GCAGAGCTGGAAGATCTTATTCAACAGACAGATTTCCAAAGACAGACAGTCAAGCCCGAGCC 360

QY 363 tccgcactgctactatgaacagatcactcactaataagaattcttcttgtagaattc 422
|||||
Db 361 TCCGCAATCGACTATGATGAACAGTATCACTCAATGAATCTATCTGATGAAATTT 420
QY 423 ataacgagagagactctgctatgcttacaataatccacttgatctcatttgagaag 482
|||||
Db 421 ATACTGAGAGGCAATCTGATGATGCTTACAAAATTCACATTTGGATCTCATTTGAGAG 480
QY 483 taccactctatgttttaagagttcttgagaagaacaa 521
|||||
Db 481 TACCACACTATGTTTAAAGGTTTCTGGAGACACGCA 519
RESULT 14
AV692032 602 bp mRNA linear EST 16-JAN-2002
LOCUS AV692032 GKC Homo sapiens cDNA clone GKACD01 5', mRNA sequence.
DEFINITION AV692032
ACCESSION AV692032
VERSION AV692032.1 GI:10293895
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 602)
XU, X., HUANG, J., XU, Z., QIAN, B., ZHU, Z., YAN, Q., CAI, T., ZHANG, X.,
XIAO, H., QU, J., LIU, F., HUANG, Q., CHENG, Z., LI, N., DU, J., HU, W.,
SHEN, K., LU, G., FU, G., ZHONG, M., XU, S., GU, W., HUANG, W., ZHAO, X.,
HU, G., GU, J., CHEN, Z., and HAN, Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1. 602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="GKACD01"
/clone_1lb="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); site_1: EcoRI; site_2:
XhoI"
BASE COUNT 165 a 133 c 120 g 182 t 2 others
ORIGIN
Query Match 32.4%; Score 509.4; DB 9; Length 602;
Best Local Similarity 98.7%; Pred. No. 1.9e-65;
Matches 513; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 gaaatctgtctgagatgaagctttgacgcttgacgtctctgtaccatgtctctc 61
|||||
Db 1 GAAATATGCTGTGGATGAAGCTTTGACGCTTGACGCTCTGTACCAATGTTCTCTT 60
QY 62 ctgtgagcagcatgtctctgctgcgtccagagtgccaaagtctcagctcttccctagaac 121
|||||
Db 61 CTGTGAGCAGCAGTCTCTCCGCTTTCAGAGTGCGCAAGTCTAGCTCTCTCTCTAGAAC 120
QY 122 ctctagcagatcagatctacagaactctactacaacatagaagatttctctgca 181
|||||
Db 121 CTCTAGGCAAGTCAAGTCTTACAGAACTTACTACAACTATGAGATGTCTCTGCA 180

QY 182 gccgtaacagctgaccltattgtgaagaanaaaacaaagtcatttttggtaattgcatc 241
|||||
Db 181 GCCGCTAAACAGCTGACCTTATTGTGAAGAAAACAAAGTCCATTTTGTGAATGATC 240
QY 242 tgatgtgacaaatgtgaaagcccatctaattgagggaaattccatgacgtctgtc 301
|||||
Db 241 TGATGTGACAAATGTGAAGCCCATTTAAAGTGAAGGGAATTCATGCACTGTCTTGGCT 300
QY 302 ggcagagctggaagatcttattcaacagcagattcccaagacacagtcagcccgagc 361
|||||
Db 301 GGCAGAGCTGGAAGATCTTATTCAACAGCAGATTTCCACACACAGTCAGCCCCGAGC 360
QY 362 ctccgcatcgtactatgaacagatcactcactaataagaattctattcttgtagaatt 421
|||||
Db 361 CTCGCAATCGACTATGATGAACAGTATCACTCAATGAATCTATCTTGATGAGAAAT 420
QY 422 tatactgagagagcattcctgatattgcttaaaaaatccacattgattccttattgaga 481
|||||
Db 421 TATACTGAGAGGCACTCTGATGATGCTTACAAAATCCACATTTGGATCTCATTTGAGAA 480
QY 482 gtaccactctatgttttaagagttcttgagaagaacaa 521
|||||
Db 481 GTACCACACTATGTTTAAAGGTTTCTGGAGACACGCA 520
RESULT 15
AV698425 621 bp mRNA linear EST 16-JAN-2002
LOCUS AV698425 GKC Homo sapiens cDNA clone GKADP03 5', mRNA sequence.
DEFINITION AV698425
ACCESSION AV698425
VERSION AV698425.1 GI:10300396
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 621)
XU, X., HUANG, J., XU, Z., QIAN, B., ZHU, Z., YAN, Q., CAI, T., ZHANG, X.,
XIAO, H., QU, J., LIU, F., HUANG, Q., CHENG, Z., LI, N., DU, J., HU, W.,
SHEN, K., LU, G., FU, G., ZHONG, M., XU, S., GU, W., HUANG, W., ZHAO, X.,
HU, G., GU, J., CHEN, Z., and HAN, Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1. 621
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="GKADP03"
/clone_1lb="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); site_1: EcoRI; site_2:
XhoI"
BASE COUNT 166 a 137 c 129 g 185 t 4 others
ORIGIN
Query Match 32.3%; Score 507.4; DB 9; Length 621;
Best Local Similarity 98.5%; Pred. No. 3.7e-65;
Matches 511; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
OY 3 aaaaatgctgttggatgaaagcttggcaagccttgacgtcttctgttaccattgttctcttc 62
   |||||||
Db 1 AAAATTGCTGTGGGATGAAAGCTTTCACAGCTTCGAGTCCCTGTACCAATTGTCCTTC 60
   |||||||
OY 63 tgttgagcaatgtcttctgcttccagagtgagccaaagttcagctgtcttccctagaacc 122
   |||||||
Db 61 ttgtgagcagcattgtcttccgcttccagagtgagccaaagttcagctgtcttccctagaacc 120
   |||||||
OY 123 tctaggcaagttcagaagttctacagaatcttactacacaatagagatgttctcttgagcag 182
   |||||||
Db 121 ttctaggcaagttcagaagttctacagaatcttactacacaatagagatgttctcttgagcag 180
   |||||||
OY 183 ccggttaacagctgaccttatttggagaagaaaaaaccaagtcatttcttggtaaagtcact 242
   |||||||
Db 181 CCGGTAACAGCTGAGCTTATGTGAAAGAAAAACAAGTCCATTTTGTGTAATGCAATCT 240
   |||||||
OY 243 gatgtcgaaatgtgaaagcccaattaaatgtgagcggaaatccatgcaagtgtctgtcg 302
   |||||||
Db 241 GATGTCGACAAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGCTG 300
   |||||||
OY 303 gcaagcgtggaaatcttactaacagcagatlttccaacgacagtlcagcccccgagcc 362
   |||||||
Db 301 GCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGCTCAGCCCCGAGCC 360
   |||||||
OY 363 tccgcatcgtaactaagaacagatcactcactaaatgaatctatcttctgtgataaatt 422
   |||||||
Db 361 TCCGCATGCTACTATGAACAGTATCATCTCAATAATGAATCTATTCTGATAGAAATT 420
   |||||||
OY 423 ataactgagagcattcccgatatagtcttaaaaaaatccacattgatactcatttgaagaag 482
   |||||||
Db 421 ATAACTGAGAGGCAATCCGATATGCTTACAAAAATCCACATTGGANCCCATTTGAGAAG 480
   |||||||
OY 483 taccacactatgttctlaaaggttctctgaaagaacaa 521
   |||||||
Db 481 TACCACACTATGNTNTAAAGTTTCTGGAAACAGCCA 519
   |||||||
```

Search completed: September 18, 2002, 16:32:00
Job time: 4203 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:43:08 ; Search time 23.83 Seconds

(without alignments)
584.937 Million cell updates/sec

Title: US-09-980-881-2

Perfect score: 1911

Sequence: 1 MKICSLAVLPVLFCEQHV.....IKYFTSNPPEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query length	ID	Description
1	692	36.2	415 1 CBPB_RAT	P19223 ratnus norv
2	641	33.5	417 1 CBPB_HUMAN	P15086 homo sapien
3	629.5	32.9	401 1 CBPB_PIG	P09955 sus scrofa
4	628.5	32.9	417 1 CBPC-MOUSE	P15089 mus musculu
5	627	32.8	416 1 CBPB-CANFA	P55261 canis fami1
6	619.5	32.4	417 1 CBPC-HUMAN	P15088 homo sapien
7	580	30.4	306 1 CBPB_BOVIN	P00732 bos taurus
8	553	28.9	417 1 CBP2_RAT	P19232 ratnus norv
9	546	28.6	309 1 CBPC_RAT	P21961 ratnus norv
10	533.5	27.9	419 1 CBPA_BOVIN	P00730 bos taurus
11	529.5	27.7	419 1 CBP1_RAT	P00731 ratnus norv
12	508	26.6	417 1 CBP2_HUMAN	P48052 homo sapien
13	489.5	25.6	419 1 CBP1_HUMAN	P15085 homo sapien
14	483	25.3	421 1 CBP2_HUMAN	O9u142 homo sapien
15	404.5	21.2	304 1 CBP2_HUMAN	P42788 stimulum vi
16	386	20.2	433 1 CBPA-ANOGA	O02350 anopheles g
17	369.5	19.3	303 1 CBPB-ASTFL	P04059 astacus flu
18	299.5	15.7	430 1 YHT-YEAST	P38836 saccharomyc
19	277.5	14.5	424 1 CBPT-THYU	P29068 thermocactin
20	271	14.2	451 1 CBPS-STREG	P18143 streptomycin
21	251	13.1	434 1 CBPS-STREG	P39041 streptomycin
22	118.5	6.2	707 1 ORC1-SCPO	P54789 schizosacch
23	105	5.5	1255 1 PER2-HUMAN	O15055 homo sapien
24	100	5.2	376 1 YQGT-BACSU	P54497 bacillus su
25	99.5	5.2	445 1 TR5H-CHICK	P70080 gallus gall
26	98.5	5.2	448 1 TY3H-ANGAN	O42091 anguilla an
27	97	5.1	663 1 TAZ1-SCPO	P09005 schizosacch
28	95	5.0	444 1 TR5H-HUMAN	P17752 homo sapien
29	93.5	4.9	3587 1 SRF1-BACSU	P27206 bacillus su
30	91.5	4.8	465 1 STHA-ECOLI	P27306 escherichia
31	91	4.8	297 1 Y103-HUMAN	O15006 homo sapien
32	91	4.8	444 1 TR5H-RABIT	P17290 cycloclagus
33	91	4.8	982 1 ENV_SFV3L	P27399 simian foam

34	90	4.7	1451	1	SPT6_YEAST	P23615 saccharomyc
35	89.5	4.7	682	1	CBA_BACUH	O86170 bacillus th
36	89	4.7	525	1	PUR9_BUCAI	P57143 b bifunctio
37	88.5	4.6	319	1	UN30_CAEEL	P52906 caenorhabdi
38	88	4.6	454	1	CBPH_LOPAM	P37892 lophus ame
39	88	4.6	1150	1	C9EA_BACTA	O92n19 bacillus th
40	87	4.6	516	1	YKE7_YEAST	P36090 saccharomyc
41	86	4.5	523	1	SYK_PRRHO	O57963 pyrococcus
42	85.5	4.5	396	1	ENP1_BACSH	O03415 bacillus sp
43	85.5	4.5	476	1	CBPH_RAT	P15087 ratnus norv
44	85	4.4	481	1	TR5H_XENLA	O92142 xenopus lae
45	85	4.4	1047	1	RIRL_CHLTR	O84834 chlamydia t

ALIGNMENTS

```

RESULT 1
CBPB_RAT ID CBPB_RAT STANDARD: PRT: 415 AA.
AC P19223;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
CPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clausen E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
family."
RL J. Biol. Chem. 263:17837-17845(1988).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M23959; AAA40872.1; JOINED.
DR EMBL: M23947; AAA40872.1; JOINED.
DR EMBL: M23950; AAA40872.1; JOINED.
DR EMBL: M23952; AAA40872.1; JOINED.
DR EMBL: M23953; AAA40872.1; JOINED.
DR EMBL: M23954; AAA40872.1; JOINED.
DR PIR: A32129; A32129.
DR HSSP: P09955; INSA.
DR MEROPS: M14.003; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypep.
DR Pfam: PF02244; Propep_M14.1.
DR Pfam: PF00246; Zn_carboxypep.1.
DR PRINTS: PR00765; CROXYPEPTASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
DR HydroLase: Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
KW SIGNAL
FT PROPEP 1 13
FT CHAIN 14 108
FT METAL 109 415
FT METAL 174 174
FT METAL 177 177
FT METAL 177 177
ZINC (BY SIMILARITY).
ZINC (BY SIMILARITY).
ZINC (BY SIMILARITY).

```

FT METAL 302 302 ZINC (BY SIMILARITY).
 FT ACT_SITE 376 376 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 171 184 BY SIMILARITY.
 FT DISULFID 243 266 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 SQ SEQUENCE 415 AA; 47515 MW; 8BA0CCADE30B672 CRC64;

Query Match 36.2%; Score 692; DB 1; Length 415;
 Best Local Similarity 39.2%; Pred. No. 6,9e-51;
 Matches 150; Conservative 61; Mismatches 126; Indels 46; Gaps 7;

QY 1 MKLSLAVIPYVIFCGQHAFVAFSGOYLALPRTSRQOYVQLNTTYYEIVLMQPVADL 60
 DB 1 MLLALVVALAHASEH---FDGNRYRVSVHGEDHVNLIQELANTKEIDFWKPDASAT 57
 QY 61 LIYKKOVHFEVNASDVNDVNAHLNVSIGIPCVSLADVEDLIQOQISNDYSPRASASY 120
 DB 58 OVKPLTVDFVYKAEVDADVADENFLEENVEVHREYVLSVNRALLESQFDSHT---RASGHSY 114
 QY 121 EYHSLNEIYSWIEETTERHPDMLTKIHGSSFEKYPYLVYKSGKQOTAKNAIWDGCI 180
 DB 115 TKYNKMETIEMWIOQVATNDPDLVTOVSIGTTEGGRMYVLKI-GKTRPKPAIFIDCGF 173
 QY 181 HAREWISPAFCIMFI-----GH-----NRMRK 203
 DB 174 HAREWISPAFCQFVRAVTVYNOELHMKOLDELDFYLVPNYIDYVYTWTKDRMRK 233
 QY 204 NRSFYANNHICIGDNLNSFNYSKHCERGASSSCSEYTCGLYPESEPEVKAASFLEARNI 263
 DB 234 TRSTMAGSSCLGVRRPNRN-NAGMCEVGAASRSPCEYCGPAPASEKETALADFITNNL 292
 QY 264 NQKAYISMSYSGHIVPEFYSYTRSKSKDHEELSLVASEVRAIDKTSKFTYTHGSE 323
 DB 293 STTKALVTHSYSQMMLYPYSYDKLPENYELNALVKGAKEL-ATLHGTKYTYGPGAT 351
 QY 324 TLXLPAGGDDWITLYDGIRKSYFT 346
 DB 352 TIYPAAGSDMSYDQCIKSYFT 374

RESULT 2
 CBBR_HUMAN
 ID CBBR_HUMAN STANDARD; PRT: 417 AA.
 AC P15086; O60834;
 DT 01-APR-1990 (rel. 14, Created)
 DT 15-DEC-1998 (rel. 37, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein) (PASP).
 GN CPB1 OR CPB OR PCPB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
 RC TISSUE-Pancreas;
 RX MEDLINE=92129345; PubMed=1370825;
 RA Yamamoto K.K., Foussette A., Chow P., Wilson H., el Shami S., French C.K.;
 RA "Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of a human serum marker for acute pancreatic procarboxypeptidase B.";
 RL J. Biol. Chem. 267:2575-2581(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Pancreas;
 RX MEDLINE=98182241; PubMed=9524066;
 RA Aloy P., Cataus L., Villegas V., Reverter D., Vendrell J., Aviles F.X.;
 RA "Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B.";

RL BIOL. Chem. 379:149-155(1998).
 RN [3]
 RP SEQUENCE OF 16-43.
 RC TISSUE-Pancreas;
 RX MEDLINE=89153096; PubMed=2920728;
 RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
 RT Purification and properties of five different forms of human procarboxypeptidases "I";
 RL Eur. J. Biochem. 179:609-616(1989).
 CC -I- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O = peptide + L-Lysine(or L-arginine).
 CC -I- TISSUE SPECIFICITY: PANCREAS.
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE ZINC CARBOXYPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@sdb.ch).
 CC -----
 DR EMBL: M61057; AAA6973.1; -;
 DR EMBL: AJ224866; CAI2163.1; -;
 DR PIR: S02812; S02812.
 DR PIR: A42332; A42332.
 DR HSSP: P09955; 1NSA.
 DR MEROPS: M14.003; -;
 DR MIM: 114852; -;
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_Carboxypept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR PRINTS: PR00765; CARBOXYPPASRA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL 1 15
 FT PROPEP 16 110
 FT CHAIN 111 417
 FT METAL 176 176
 FT METAL 179 179
 FT METAL 304 304
 FT ACT_SITE 378 378
 FT DISULFID 173 186
 FT DISULFID 245 268
 FT DISULFID 259 273
 FT CONFLICT 16 16
 FT CONFLICT 17 17
 FT CONFLICT 37 37
 FT CONFLICT 208 208
 FT CONFLICT 245 245
 SQ SEQUENCE 417 AA; 47366 MW; BBIQF212D830305E CRC64;

Query Match 33.5%; Score 641; DB 1; Length 417;
 Best Local Similarity 35.3%; Pred. No. 1.4e-46;
 Matches 135; Conservative 71; Mismatches 132; Indels 44; Gaps 7;

QY 3 LCSLAVIPYVIFCGQHAFVAFSGOYLALPRTSRQOYVQLNTTYYEIVLMQPVADL 61
 DB 1 MLLALVVALAHASEHGEDHGEKRVVNEEDENHINIIRELASTQIDFWKPDVSQ 60
 QY 62 IYKKOVHFEVNASDVNDVNAHLNVSIGIPCVSLADVEDLIQOQISNDYSPRASASYE 121
 DB 61 IKPHSTVDFVYKAEPTVYEVNVLKQNELQYVLSNLNVEADPSRV---RATGHSYE 117
 QY 122 EYHSLNEIYSWIEETTERHPDMLTKIHGSSFEKYPYLVYKSGKQOTAKNAIWDGCIH 181
 DB 118 KYNKMETIEMWIOQVATNDPDLVTOVSIGTTEGGRMYVLKI-GKAKQNKRAIIMDGGFH 176
 QY 182 AREWISPAFCIMFI-----GH-----NRMRKN 204

Db	177	AREWISPAFCQMFBEAVRTYGRIGQYTELLNKLDFFVLYVLNIDGTYITWTSNRFRKRT	236
Qy	205	RSFYANNHCIGCTDINSNFVSKHMCCEGASSSSCSEYTCGLYPESEPEYKAVASFLRRNIN	264
Db	237	RSTHTGSSCICCTDPPNRRNF-DAGMCEIGASRNPCDETCYCGPAASESEKETFKLADPIRKLIS	295
Qy	265	QIKAVISMHSQHVHIPPYSTTRKSKSDHELSLVASEAVRALDIKTSKNTRYTHGHQSET	324
Db	296	SIRKAVILTHSYQMMKLYPSYAVYKIGENNAELNALNATVKEI-ASLHGKRYTYGGPCTT	354
Qy	325	LYLAPGGDDWIVDGLIKRYSPT	346
Db	355	ITYPAAGSGSDMAYTDOGITRYSFT	376
RESULT 3			
CPBP_PIG	STANDARD;	PRT;	401 AA.
AC	P09955;		
DT	01-MAR-1989 (Rel. 10, Created)		
DT	15-JUL-1998 (Rel. 36, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Carboxypeptidase B precursor (EC 3.4.17.2).		
GN	CPB.		
OS	Sus. scrofa (Pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.		
OX	NCBI_Taxid=9823;		
MN	[1]		
RP	SEQUENCE OF 1-103.		
RX	MEDLINE=91208150; PubMed=2018774;		
RA	Burgos F.J., Salva M., Villagas V., Soriano F., Mendez E.,		
RA	Aviles F.X.;		
RT	"Analysis of the activation process of porcine procarboxypeptidase B		
RT	and determination of the sequence of its activation segment.";		
RL	Biochemistry 30:4082-4089(1991).		
RN	[2]		
RP	PRELIMINARY SEQUENCE OF 1-38.		
RX	MEDLINE=95279427; PubMed=4026847;		
RA	Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;		
RT	"Sequential homologs between procarboxypeptidases A and B from		
RT	porcine pancreas.";		
RL	Biochem. Biophys. Res. Commun. 130:97-103(1985).		
RN	[3]		
RP	X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).		
RX	MEDLINE=9114690; PubMed=1989878;		
RA	Coll M., Guasch A., Aviles F.X., Huber R.;		
RT	"Three-dimensional structure of porcine procarboxypeptidase B: a		
RT	structural basis of its inactivity.";		
RL	EMBO J. 10:1-9(1991).		
RN	[4]		
RP	STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.		
RX	MEDLINE=91027767; PubMed=2223783;		
RA	Vendrell J., Wider G., Aviles F.X., Wuehrich K.;		
RT	"Sequence-specific 1H NMR assignments and determination of the		
RT	secondary structure for the activation domain isolated from		
RT	pancreatic procarboxypeptidase B.";		
RL	Biochemistry 29:7515-7522(1990).		
RN	[5]		
RP	STRUCTURE BY NMR OF ACTIVATION PEPTIDE.		
RX	MEDLINE=9114693; PubMed=1989879;		
RA	Vendrell J., Billeter M., Wider G., Aviles F.X., Wuehrich K.;		
RT	"The NMR structure of the activation domain isolated from porcine		
RT	procarboxypeptidase B.";		
RL	EMBO J. 10:11-15(1991).		
RN	[6]		
RP	STRUCTURE BY NMR OF ACTIVATION PEPTIDE.		
RX	MEDLINE=93044373; PubMed=1422143;		
RA	Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,		
RA	Huber R., Wuehrich K.;		
RT	"Comparison of the NMR solution structure with the x-ray crystal		
RT	structure of the activation domain from procarboxypeptidase B.";		

BL	J. Biomol. NMR 2:1-10(1992).	
CC	-1- CATALYTIC ACTIVITY: peptidyl-L-Lysine(or L-arginine) + H(2)O =	
CC	peptide + L-Lysine(or L-arginine).	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE	
CC	ZINC CARBOXYPEPTIDASE FAMILY.	
CC	-1- DATABASE: NAME=worthington enzyme manual;	
CC	WWW=http://www.worthington-biochem.com/manual/c/COB.html".	
DR	PIR; B29181; B29181.	
DR	PDB; 1PBA; 31-OCT-93.	
DR	PDB; INSA; 24-DEC-97.	
DR	MEROPS; M14.003; "	
DR	InterPro: IPR003146; Proped.M14.	
DR	InterPro: IPR000834; zn_carboxpept.	
DR	Pfam; PF02244; Propep_M14; 1.	
DR	Pfam; PF00246; zn_carboxpept; 1.	
DR	PRINTS; PR00765; CARBOXYPTASE.	
DR	PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.	
DR	PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.	
KW	Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;	
KW	3D-structure.	
FT	PROPEP	1 95
FT	CHAIN	96 401
FT	DISULFID	158 171
FT	DISULFID	230 253
FT	DISULFID	244 258
FT	METAL	161 161
FT	METAL	164 164
FT	METAL	289 289
FT	METAL	341 341
FT	ACT_SITE	363 363
FT	ACT_SITE	363 363
FT	STRAND	14 17
FT	HELIX	20 31
FT	HELIX	32 32
FT	HELIX	43 45
FT	HELIX	50 52
FT	HELIX	61 69
FT	TURN	70 72
FT	TURN	75 76
FT	STRAND	75 76
FT	SEQUENCE	401 AA; 45713 MW; 53129AFL159A26348 CRC64;

[illegible]

```

RESULT 4
CBPC_MOUSE
ID CBPC_MOUSE STANDARD: PRT: 417 AA.
AC P15089;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90062123; PubMed=2584208;
RA Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA Serafin W.E.;
RT "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT novel member of the carboxypeptidase gene family."
RL J. Biol. Chem. 264:20094-20099(1989).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O -> peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J05118; AAA37369.1; -
DR PIR: A34487; A34487.
DR HSSP: P09955; 1NSA.
DR MEROPS: M14.010; -.
DR MGD: MGI:88479; CPA3.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE.
FT CHAIN 110 417 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 245 268 BY SIMILARITY.
SQ SEQUENCE 417 AA; 48790 MW; AAB300A068D1BA6D CRC64;

Query Match 33.9%; Score 628.5; DB 1; Length 417;
Best Local Similarity 37.0%; Pred. No. 1.5e-45;
Matches 140; Conservative 60; Mismatches 131; Indels 47; Gaps 8;
OY 6 LAVLPVILFCEQHFARQSGOVLALPRTSRQOVLONTTTEIVLMQPTADLIYKK 65
Dy 7 MAVIYTLAIAPVH---FDREKVFVKLQNEKHSVYLNKLGSLQSLDWMYDAIHDAVNV 63
OY 66 KQVIEFVNASVDNKKALNNSGIPCSVTLADVEDLIQOQIS-NDYVSPRASASVEQYH 124
Dy 64 MTPVFRVSEKESQTIQSLTEQHKIHYELLHDLQELKEKQFDVDELAGRHS---YAKYN 120

```

```

OY 125 SLNEIYSWIEFIERHPDMLTKIHIGSFPEKYPYLAKYSGEQAKNAIWDGCIHARE 184
Dy 121 DMKRTYSWTEKMLEKHPDENVSRIKISGTVEDNPLVYKTI-GRKDGERRAIFMDCGHARE 179
OY 185 WISPAFCLMEI-----GHN-----RWMRKRSF 207
Dy 180 WISPAFCQFWEYQATKSYGNKIMTKLIDRMNRYVLPVENVGDIYIMSWTQDMKRRNSR 239
OY 208 YANNHICIGDLDNNEFYSKRWCEGASSSCSEFYCGLYESEPVEKAVASFLRRINQIK 267
Dy 240 NONSTICIGDLDNNEF-VYSWDSPPNTNKPCLNVRGPPAPESEKETAAYNFTIRSLNSIK 298
OY 268 AVISMHSYQHIYFYPYSYRSKSKDHEELSLVASEAVRIDKTSNRTYTHGSGSETLYL 327
Dy 299 AVTFHSYSQMLIPYGYFKLPENQDILKVARIAITDAL-STRETRITLYGPIASTLYK 357
OY 328 APEGGDWYIDGIRKSF 345
Dy 358 TSGSSLDWYDGLIKHTF 375

RESULT 5
CBPB_CANPA
ID CBPB_CANPA STANDARD: PRT: 416 AA.
AC P55261;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE membrane associated protein) (ZAP47).
GN CPB1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Fukuko S.-I.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-lysine(or L-arginine) + H(2)O -
CC peptide + L-lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D78348; BAAL1366.1; -
DR HSSP: P09955; 1PBA.
DR MEROPS: M14.003; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 110 416 CARBOXYPEPTIDASE B.
FT METAL 175 175 ZINC (BY SIMILARITY).
FT METAL 178 178 ZINC (BY SIMILARITY).
FT METAL 303 303 ZINC (BY SIMILARITY).
FT ACT_SITE 377 377 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 244 267 BY SIMILARITY.

```



```

Db 177 ARHWISAPFQWVYQATKTYGKRNKTKILDRNFIILPEVNDGTYMSWTKNRMRKN 236
QY 205 RSYFANNHCIGTDLNNSNFVSKHMCCEGASSSCSEYTCGLYPESEPEVKAVASFLRNIN 264
Db 237 RSKNNSKICIGTDLNRF-NASWNSIPNTNDPCADVNRGAPSEKREKAVTFINSHLN 295
QY 265 QIKAYISMHSYSHIYFPYSYTRSKSDHELSLVASEAVRAIDKSKNTRYHGHGSET 324
Db 296 ELKRYTFHSHYSQMLFPYGYTSKLPNHEDLAKVAKIGTDVL-STRYETRYIYGIEST 354
QY 325 LYLAPGCGDDWIDYLGIKYSF 345
Db 355 IYPISSSIDMAYDGIKHTF 375

```

```

RESULT 7
CBP2_BOVIN STANDARD; PRT: 306 AA.
AC P00732:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B (EC 3.4.17.2).
GN CPB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN 11
RP MEDLINE=75217824; PubMed=1057162;
RA Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT "Amino-acid sequence of bovine carboxypeptidase B.";
RT Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
RN 12
RP SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
RA MEDLINE=74260705; PubMed=4833744;
RA Schmidt J.J., Hirs C.H.W.;
RT "Primary structure of bovine carboxypeptidase B. Inferences from the
RT locations of the half-cystines and identification of the active site
RT arginine.";
RT J. Biol. Chem. 249:3756-3764(1974).
RN 13
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RA MEDLINE=78265065; PubMed=957425;
RA Schmidt M.F., Herriott J.R.;
RT "Structure of carboxypeptidase B at 2.8-A resolution.";
RT J. Mol. Biol. 103:175-190(1976).
RN 14
RP ACTIVE SITE.
RA MEDLINE=70007159; PubMed=5344132;
RA Plummer T.H. Jr.;
RT "Isolation and sequence of peptides at the active center of bovine
RT carboxypeptidase B.";
RT J. Biol. Chem. 244:5246-5253(1969).
RN 15
RP ACTIVE SITE.
RA MEDLINE=73061487; PubMed=4565668;
RA Kimmel M.T., Plummer T.H. Jr.;
RT "Identification of a glutamic acid at the active center of bovine
RT carboxypeptidase B.";
RT J. Biol. Chem. 247:7864-7869(1972).
RN 16
RP CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
RN peptide + L-Lysine(or L-arginine).
RN -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
RN ZINC CARBOXYPEPTIDASE FAMILY.
CC PIR: A00912; CPBOB.
CC PDB: 1CPB; 30-SEP-83.
CC MEROPS: M14.003; -.
CC InterPro: IPR000834; Zn_carboxypept.
CC Pfam: PF00246; Zn_carboxypept; 1.

```

```

DR PRINTS; P00765; CRBOXYPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT DISULFID 63 76
FT DISULFID 135 158
FT DISULFID 149 163
FT METAL 66 66
FT METAL 69 69
FT METAL 194 194
FT ACT_SITE 246 246
FT ACT_SITE 268 268
SQ SEQUENCE 306 AA; 34612 MW; C329D2655C4A172 CRC64;
NUCLEOPHILE.

```

```

Query Match 30.4%; Score 580; DB 1; Length 306;
Best Local Similarity 43.6%; Pred. No. 1,2e-41;
Matches 115; Conservative 44; Mismatches 65; Indels 40; Gaps 5;

```

```

QY 120 YEOYSLNMEIYSEWIEFTRHPDMLTKHIGSSFERYPLVYLVGSGKQTAKNALWIDOG 179
Db 6 YEKYNWETLEAMTEQVASENDLISRAIGTFLGNTIYLKV-GKGSNKPAYFMDCG 64
QY 180 IAREWISPAFLMPT-----GH-----NRMR 202
Db 65 FAREWISPAFQWVYREAVRTYGREIHMTFELDKLDFVILPVVNDIYITWTNRMWR 124
QY 203 KNRSEFANNHCIGTDLNNSNFVSKHMCCEGASSSCSEYTCGLYPESEPEVKAVASFLRN 262
Db 125 KTRSTRAGSCCTGTDLNR-F-DAGMCSIGASNNPCSEYTCGSAASEKESKAVADFIRNH 183
QY 263 INQIKAYISMHSYSHIYFPYSYTRSKSDHELSLVASEAVRAIDKSKNTRYHGHGS 322
Db 184 ISSIRAYLIHSHYSQMLFPYGYTSKLPNNVELMTLAKAVKKI-ASLHGTTISYGPGA 242
QY 323 ELYLAPGCGDDWIDYLGIKYSF 346
Db 243 TTIYPAAGSDMAYDGIKYSF 266

```

RESULT 8

```

CBP2_RAT STANDARD; PRT: 417 AA.
AC P19222:
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN CPA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA MEDLINE=89034323; PubMed=3182871;
RA Gardell S.J., Craik C.S., Clauser E., Goldsmith E.J., Stewart C.B.,
RA Graf M., Rutter W.D.;
RT "A novel rat carboxypeptidase, CPA2: characterization, molecular
RT cloning, and evolutionary implications on substrate specificity in
RT the carboxypeptidase gene family.";
RT J. Biol. Chem. 263:117828-117836(1988).
RN 12
RP SEQUENCE OF 131-143 FROM N.A.
RA MEDLINE=95386501; PubMed=7657630;
RA Normant E., Gros C., Schwartz J.C.;
RT "Carboxypeptidase A isoforms produced by distinct genes or
RT alternative splicing in brain and other extrapancreatic tissues.";
RT J. Biol. Chem. 270:20543-20549(1995).
RN 13
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA MEDLINE=92105124; PubMed=1761558;

```


QY 263 INQIATYISMHSYCHIVPEYSTRSKSKDEELSLVASEAVRAIDKTSKRTYTHGHGS 322
 Db 186 LNSIKAVYTFHSYSQMLFPYGYTIKLEPNQDLKVARIRTDVL-SSRYETRIYIGPIA 244
 QY 323 ETLVAPGGGDWYDGIKYSF 345
 Db 245 STIKTSGSSLDWADGIRKTF 267

RESULT 10
 CBPA_BOVIN
 ID CBPA_BOVIN STANDARD; PRT; 419 AA.
 AC P00730;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase A precursor (EC 3.4.17.1).
 GN CPA.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91151335; PubMed=1998496;
 RT "Cloning and nucleotide sequence of a bovine pancreatic
 RT preprocarboxypeptidase A cDNA."
 RL Biochem. Biophys. Res. Commun. 175:110-116(1991).
 RP [2]
 RC SEQUENCE FROM N.A.
 RX TISSUE=Pancreas;
 RA MEDLINE=96096549; PubMed=8522204;
 RT "Cloning, sequencing and expression of the gene encoding a major
 RT allotypic preprocarboxypeptidase A from bovine pancreas."
 RL Gene 165:333-334(1995).
 RN [3]
 RP SEQUENCE OF 111-417.
 RA MEDLINE=71155183; PubMed=5102489;
 RT Bradshaw R.A., Walsh K.A., Neurath H.;
 RT "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
 RT chymotryptic peptides of the cyanogen bromide fragment F-I."
 RL Biochemistry 10:938-950(1971).
 RN [4]
 RP REVISIONS TO 138 AND 141.
 RA MEDLINE=72138789; PubMed=5143102;
 RT Petra P.H., Hermanson M.A., Walsh K.A., Neurath H.;
 RT "Characterization of bovine carboxypeptidase A (Allan)."
 RL Biochemistry 10:4023-4025(1971).
 RN [5]
 RP SEQUENCE OF 17-120.
 RA MEDLINE=89150306; PubMed=3147705;
 RT Wade R.D., Haas G.M., Kumar S., Walsh K.A., Neurath H.;
 RT "The amino acid sequence of the activation peptide of bovine pro-
 RT carboxypeptidase A."
 RL Biochimie 70:1137-1142(1988).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
 RA MEDLINE=83294519; PubMed=6887246;
 RT Rees D.C., Lewis M., Lipscomb W.N.;
 RT "Refined crystal structure of carboxypeptidase A at 1.54-A
 RT resolution."
 RL J. Mol. Biol. 168:367-387(1983).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
 RA MEDLINE=96003618; PubMed=7556081;
 RT Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
 RT "The three-dimensional structure of the native ternary complex of
 RT bovine pancreatic procarboxypeptidase A with proproteinase E and
 RT chymotrypsinogen C.";

RL EMBL J. 14:4387-4394(1995).
 RN [8]
 RP VARIANT ALLELIC.
 RA MEDLINE=69283620; PubMed=5817619;
 RT Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
 RT "Identification of the amino acid replacements characterizing the
 RT allotypic forms of bovine carboxypeptidase A."
 RL Biochemistry 8:2762-2768(1969).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
 CC amino acid.
 CC -1- SUBUNIT: MONOMER. THE ZYMOMEN IS SECRETED AS A TERNARY COMPLEX
 CC COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
 CC PROPROTEINASE E.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -1- DATABASE: NAME=Worthington enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/C/COA.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M61851; AAA30426.1; -;
 DR EMBL: M61851; AAA30427.1; -;
 DR EMBL: Z33906; CAA83955.1; -;
 DR PIR: JN0126; CPBOA.
 DR PIR: A31406; A31406.
 DR PDB: 3CPA; 15-JAN-87.
 DR PDB: 4CPA; 22-OCT-84.
 DR PDB: 5CPA; 15-JAN-87.
 DR PDB: 6CPA; 15-OCT-91.
 DR PDB: 7CPA; 31-JAN-94.
 DR PDB: 8CPA; 31-JAN-94.
 DR PDB: ICBP; 31-JAN-94.
 DR PDB: ICBP; 15-OCT-94.
 DR PDB: 2CTB; 31-JAN-94.
 DR PDB: 2CTB; 31-JAN-94.
 DR PDB: 1ART; 01-AUG-96.
 DR PDB: 1ARM; 17-AUG-96.
 DR PDB: 1BAV; 01-APR-97.
 DR PDB: 1YME; 12-FEB-97.
 DR PDB: 1CPX; 05-AUG-98.
 DR PDB: 1PYT; 27-JAN-97.
 DR MEROPS: M14.001; -;
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxypept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxypept; 1.
 DR PRINTS: PR00765; CRBOXYPEPTSEA.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
 KW 3d-structure; Polymorphism.
 FT SIGNAL 1 16
 FT PROPEP 17 110
 FT CHAIN 111 419
 FT METAL 179 179
 FT METAL 182 182
 FT METAL 306 306
 FT ACT_SITE 358 358
 FT ACT_SITE 380 380
 FT DISULFID 248 271
 FT VARIANT 289 289
 FT VARIANT 338 338
 FT VARIANT 415 415
 FT CONFLICT 95 415
 FT CONFLICT 199 199
 FT CONFLICT 203 203
 FT CONFLICT 224 224
 FT CONFLICT 224 224

ACTIVATION PEPTIDE.
 CARBOXYPEPTIDASE A.
 ZINC.
 ZINC.
 ZINC.
 PROTON DONOR.
 NUCLEOPHILE.
 I -> V (IN ALLELIC VARIANT).
 E -> A (IN ALLELIC VARIANT).
 L -> V (IN ALLELIC VARIANT).
 S -> L (IN REF. 5).
 D -> N (IN REF. 3).
 D -> N (IN REF. 3).
 D -> N (IN REF. 3).

```

FT CONFLICT 232 232 0 -> E (IN REF. 3).
FT CONFLICT 295 295 D -> N (IN REF. 3).
FT TURN 114 116
FT TURN 119 120
FT HELIX 125 138
FT TURN 140 142
FT STRAND 143 150
FT TURN 152 153
FT STRAND 156 162
FT STRAND 171 176
FT TURN 180 181
FT HELIX 183 199
FT TURN 200 202
FT HELIX 204 212
FT STRAND 214 218
FT HELIX 223 231
FT TURN 232 232
FT TURN 234 235
FT STRAND 242 242
FT TURN 244 245
FT STRAND 249 249
FT TURN 253 254
FT TURN 261 262
FT STRAND 267 267
FT TURN 270 271
FT TURN 273 274
FT STRAND 275 275
FT TURN 280 281
FT HELIX 284 296
FT STRAND 299 307
FT STRAND 311 314
FT TURN 324 325
FT HELIX 326 344
FT STRAND 349 352
FT HELIX 353 356
FT TURN 357 357
FT HELIX 364 370
FT TURN 371 372
FT STRAND 388 389
FT TURN 393 416
SQ SEQUENCE 419 AA; 47082 MW; 21886407B3BFC452 CRC64;

Query Match 27.9%; Score 533.5; DB 1; Length 419;
Best Local Similarity 34.3%; Pred. No. 1.6e-37;
Matches 134; Conservative 61; Mismatches 133; Indels 63; Gaps 11;

OY 3 LCSLAVLVVPLVLFCEBOHVRPFGSGOVLALPRTSRQOVOLNLTYYEIVL-W---QP 56
DB 4 LLLSVLLGALGKED---FVGHQVLRITADEAEVQVKELEDEHHLQDLFWRGPGQP 59
OY 57 VFADLVYKKQVHFVNASDVNVKALNVSGIPCSVLADVDLI---OOOISNDTVSP 113
DB 60 -----GSPIDVRVPFBSLQAVKVFLEHAGRIKRYRMIDVOSLDEEDQHPAFASRA 111
OY 114 RASASY-YEOYHSINLETYSNIEFTTERHPDMLTKIHGSSFEKYPVLYLVKVSKEQJAK 172
DB 112 RSTNTFNAYATHTLDEIDYMDLVAEHQVLEKLGRTYEGRPVYLVFS-TGGSNRP 170
OY 173 AIVIDGIAHREWISPAFLCMFLT----- 195
DB 171 AIVIDGISHREWITQATGWFMAKKTEDYGDPSTAILDSMDIFLEIYTNPDGAFTH 230
OY 196 GHRMRMRKNSFYANNHCIGTDLNSNFVSKHMCSESSSSCSEYCYGLPESPEYKAV 255
DB 231 SQNRMRKRTSVSSSLCYVDANRNN-DAGFGKAGASSPCSEYTGKANKANSEVEYKSI 289
OY 256 ASFLRRNINQKAYISMHSYQHIIVPEYSTRKSKDHEELSLVASEAVRAIDKTSKNT 315
DB 290 VDEVKDHGN-FKAFLSIHSYSQLLVPGYGTQSIPDKTELNQAKSAVAL-KSLYGT 347
OY 316 YTHGSGSEITLYLAPGGDDMIYDLGIKYSFT 346

```

```

DB 348 YKTSITITTYQASGSDIMSYNGIKYSFT 378

RESULT 11
ID CBPL_RAT STANDARD; PRT; 419 AA.
AC P00731;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN CPAL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82105986; PubMed=6275388;
RA Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standring D.N.,
RA Pictet R.L., Valenzuela P., Rutter W.J.;
RT "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
RT characterization of the gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clausen E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family.";
RL J. Biol. Chem. 263:17837-17845(1988).
CC -I- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O -> peptide + L-
CC amino acid.
CC -I- SUBUNIT: MONOMER.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01232; CAA24542.1; -
DR EMBL: J00713; AAA40893.1; -
DR EMBL: M23990; AAA40955.1; -
DR EMBL: M23960; AAA40955.1; JOINED.
DR EMBL: M23985; AAA40955.1; JOINED.
DR EMBL: M23986; AAA40955.1; JOINED.
DR EMBL: M23987; AAA40955.1; JOINED.
DR EMBL: M23988; AAA40955.1; JOINED.
DR EMBL: M23989; AAA40955.1; JOINED.
DR PIR: A00911; CPRTA.
DR PIR: B32129; B32129.
DR HSSP: P00730; 1PVT.
DR MEROPS: M14.001; -
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept; 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 110 ACTIVATION PEPTIDE.
FT CHAIN 111 419 CARBOXYPEPTIDASE A1.
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 182 182 ZINC (BY SIMILARITY).

```

```

FT METAL 306 306 ZINC (BY SIMILARITY).
FT ACT_SITE 358 358 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 380 380 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 248 271 BY SIMILARITY.
FT CONFLICT 196 196 I -> V (IN REF. 1).
FT CONFLICT 261 263 FGM -> LGR (IN REF. 1).
FT CONFLICT 347 347 K -> E (IN REF. 1).
SQ SEQUENCE 419 AA; 47197 MW; BB002DICB9B7491 CRC64;

Query Match 27.7%; Score 529.5; DB 1; Length 419;
Best Local Similarity 34.5%; Pred. No. 3.4e-37;
Matches 134; Conservative 63; Mismatches 136; Indels 55; Gaps 10;

QY 2 KLSGLAVPIVLFCEGHVFAFGSGVLAALPRTSRQVQLQNLTYTEIVL--WQPVTA 59
DB 3 RLILSLLEAVCGNEN---FVGHQVLRIASDAEQVKKELEDEHLQDFWRDAR 58
QY 60 DLIVKKQVHEFNASDVNVKALNVSIGPCVLAADVDEL---QQQISNDTVSPRAS 116
DB 59 AGIPIDVRVPF---PSIQSVKAFLEYHGISTYIMIEDVQLLDERKQOKMSAFQARALST 114
QY 117 ASY-YEQVHSLNFIYSWIEFETERHPDMLKIHIGSSFEEKYPLVYLKSGEQATKANM 175
DB 115 DSENYATYHTLDELYERMDLVAEHQVLSKIQIGMTFEGRPILHVKFS-TGGINPRATM 173
QY 176 IDGCIHAREKISPAFCIMFIGH-----N 198
DB 174 IDGCIHAREKISPAFCIMFIGH-----N 198
QY 199 RMWRKRSEFVANNHCIGTDLNSNFVSKHMCCEGASSSCSEYCGLYPSEPEVKAVASP 258
DB 234 RMWRKRSHRHSQSLCYGVDPNRNW-DAGCGMACASSNPCEIRGFPNPEVEVKAVIYP 292
QY 259 LRRNIQIAVYSMSHSQHIYPPVYSYTRSKSDHEELSLVASEAVRAIDKTSKNTRYTH 318
DB 293 VYSHGN-IRAFISHSYSLILPYGYTSEPADQAELOQAKSAVTAL-TSLHGKFKY 350
QY 319 GHGSEFLVYLAPOGCDWYIDLGIKYSFT 346
DB 351 GSIIIDTYQASGSTIDWYSQGIKYSFT 378

RESULT 12
CPB2_HUMAN
ID CPB2_HUMAN STANDARD; PRT; 417 AA.
AC P48052;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN CPA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.
RC TISSUE=Pancreas;
RX MEDLINE=95204457; PubMed=7896805;
RA Catasus L., Vendrell J., Aviles F.X., Carreira S., Pulgarer A.,
RA Billeter M.;
RT "The sequence and conformation of human pancreatic
RT procarboxypeptidase A2. cDNA cloning, sequence analysis, and
RT three-dimensional model.";
RL J. Biol. Chem. 270:6651-6657(1995).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98111000; PubMed=9450539;
RA Reverter D., Garcia-Saez I., Catasus L., Vendrell J., Coll M.,
RA Aviles F.X.;
RT "Characterisation and preliminary X-ray diffraction analysis of human
RT pancreatic procarboxypeptidase A2.";

```

```

RL FEBS Lett. 420:7-10(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98046021; PubMed=9384570;
RA Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT "The three-dimensional structure of human procarboxypeptidase A2.
RT Deciphering the basis of the inhibition, activation and intrinsic
RT activity of the zymogen.";
RL EMBO J. 16:6906-6913(1997).
CC -I- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC residues.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U19977; AAA74425.1; -
DR PDB: IAYE; 13-JAN-99.
DR MEROPS: M14.002; -.
DR MIM: 600688; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR00834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CARBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
DR KX Hydrolyase: Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW 3D-structure.
FT SIGNAL 1 16
FT PROPEP 17 112 POTENTIAL.
FT CHAIN 113 417 ACTIVATION PEPTIDE.
FT METAL 177 177 CARBOXYPEPTIDASE A2.
FT METAL 180 180 ZINC.
FT METAL 304 304 ZINC.
FT ACT_SITE 378 378
FT DISULFID 246 269 NUCLEOPHILE.
FT DISULFID 318 352
SQ SEQUENCE 417 AA; 46840 MW; 6D8107330A2DBC58 CRC64;

Query Match 26.6%; Score 508; DB 1; Length 417;
Best Local Similarity 32.5%; Pred. No. 2.2e-35;
Matches 128; Conservative 63; Mismatches 137; Indels 66; Gaps 11;

QY 19 HVE---AFQSGVLAALPRTSRQVQLQNLTYTEIVL--WQPVTAIDIVKKQVHEFVN 73
DB 13 HIYCLETVGQVLEIYVSNEDQINKLQDAQOEHLDQDFWKSPT---TPGETAHVRVP 68
QY 74 ASDVDNVKALNVSIGPCVLAADVDELQQQISNDTVSPRAS---YYEQVSLNFIY 130
DB 69 FVNVQAVKVFLESQGIAYSIMIEDVQLDKENEMLEFNRRRSGNPNFCAVITLBEIS 128
QY 131 SWIEFETRRHDMKIKHIGSSFEEKYPLVYLKVS-GEQTKAKNAIWDGCIHAREWISPA 189
DB 129 QEMDQVLAHREGLVSKVNISSFPNRMNVLKFTSGD---KPAIMIDAGCIHAREWYTA 185
QY 190 FCLWFIQ-----HNRMWRKRSEFVANNH 212
DB 186 TALMTANKIVSDYGRKDSITSLDALDIFLLPYNPGIYFSGTKRMWRKRSTKSVSGL 245
QY 213 CIGTDLNSNFVSKHMCCEGASSSCSEYCGLYPSEPEVKAVASFLRRNIQIAVISM 272
DB 246 CVGVDPNRNW-DAGFGGPGGASSNPSCSDSYHGSPANSSEVEVKSIYDFIKSH-GVKKAFIIL 303
QY 273 HSYSQHIYFPYSYTRSKSDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETIYLAPOGG 332

```

```

FT DISUPID 248 271 BY SIMILARITY.
SQ SEQUENCE 419 AA; 47140 MW; 439FAFFFAEB58B1 CRC64;

Query Match 25.6%; Score 489.5; DB 1; Length 419;
Best Local Similarity 32.7%; Pred. No. 8e-24;
Matches 131; Conservative 60; Mismatches 127; Indels 83; Gaps 13;

QY 3 LCSAVLPVPIYLFCEQHFVAFQSGCVLALPRTSRQOVYLONLTTYEIVL--NO----- 55
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 4 LLVLSVLLGAFFGKED-----FVGHQVLRISVADEAOVKKELEDLEHQLDLFWRGPAHP 59
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 56 --PTWADIYAKKQYHFVFNASDVQNVKAHLNVSPICSVLLADEDI---QQOI----- 106
Db 60 GSPIDV-----RPF---PSIQAVKIFELSHSISTETMIEDVQSLDEEQOMFAFR 108
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 107 ----SNDVSPRASASYYEQYHSLNEIKYSWTEFIERHPDMLTKRHISSEFEKPLYLK 162
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 109 SRASSTDFFN-----YATHTLEIYDFDLVAENPHLYSKIQIGNTYEGRTIYVLK 161
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 163 VSGKEQTKAKNIWIDCGIHAREWISPAFCWFI----- 195
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 162 FS-TGSGKRPAIMIDTGHISREMYTQAGSVWFAKKITQDYGQDAFAITLDTLDFLEIV 220
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 196 -----GHNRMKRNRSFYANNHCITGIDLNSNPFYSKMCCEGASSCSSEYTCGLY 245
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 221 TNPDGFATHTSTNRMRKTRTSHTAGSLCIGDPNNM--DAGFGLSGASNSCSEYHGKFF 279
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 246 PESEFEYVAVASFLRNINQIKAVYSMSYSOHIYEPYSYGRSKSDHEELSLVASEAVR 305
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 280 ANSEYEVASIVDFVAKDHQEN-IKAFISHYSGLMLPYGYTTEPPVDDDELQLSKAANT 338
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 306 AIDKTSKNTRYTHGSESTLYLAPGSGDDWYIDLGIRKSYFT 346
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 339 AL-ASLYCTKFNYSGIKAIYQASGSTIDWYYSQGIKSYFT 378

RESULT 14
CBP4_HUMAN STANDARD; PRT; 421 AA.
ID CBP4_HUMAN
AC 09U142:
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).
GN CP44 OR CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
OX NCBI_TaxID=9606;
   [1]
RX MEDLINE=99310537; PubMed=10383164;
RA Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.I.;
RT "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT deacetylase inhibitors during differentiation of prostate epithelial
RT cancer cells."
RL Cancer Res. 59;2981-2988(1999).
CC -!- FUNCTION: Could be involved in the histone hyperacetylation
CC pathway.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF095719; AAF23230.1; -.
CC HSSP; P48052.1; IAVE.
CC -----

```

DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxpept.
 DR Pfam: PF00244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxpept; 1.
 DR PRINTS: PR00765; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_2; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolyase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
 FT SIGNAL 1
 FT PROPEP 17 113 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 114 421 CARBOXYPEPTIDASE A4.
 FT METAL 181 181 ZINC (BY SIMILARITY).
 FT METAL 184 184 ZINC (BY SIMILARITY).
 FT METAL 308 308 ZINC (BY SIMILARITY).
 FT ACT_SITE 382 382 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 250 273 BY SIMILARITY.
 SQ SEQUENCE 421 AA: 47379 MW: 918832587056C6A CRC64;

Query Match 25.3%; Score 483; DB 1; Length 421;
 Best Local Similarity 32.0%; Pred. No. 2.8e-33;
 Matches 119; Conservative 63; Mismatches 130; Indels 60; Gaps 10;

DB 76 DVDNKAHLNVSGIPCSVLLADVEDLI--QQQISNDVSPRASASY-YEOYHSINEIYS 131
 QY 23 FQSGCVLALPPTSPQVOVLOLMTTTEYL--WQPVTA-----DLIYKKQVHEFVNAS 75
 DB 21 FFGDVLRLINAVNGDEISRLSQLVNSNMKLNFMKSPSSFNRPVVLVPSVSLQAF---- 76
 QY 77 -----KSPFRSGCLEAVATIEDLOALINDNEDMOHNEGERSNNFNCAHSLAITH 131
 DB 132 WIEFTEHPDMLTKIHIGSSFEKYPPLVLYKSGKEQTAKNAIWDICGIIHANEWISPAFC 191
 QY 132 EMDNIAADPPDLARVKIKHISFENRPMYLYKESTGKGVRRPAVWLNAGIHSSEWISQAATA 191
 DB 132 LM-----FI-----GHNRMRNRSEFYANNHCT 214
 QY 192 IWTARKIVSDYORDPAITSILEKMDIFLIPVANPDGYVYTQONRLMRTSRNREGSCT 251
 DB 215 GTDLNSNFVSKHMCESGSSSCSEYTCGLYPESEPEVAVASFLRRNINQIKAYISMHS 274
 QY 252 GADPRNRW-NASFAKGASADNCPSEYHGHANSEVEVASYVDIQKHGN-FKGFIDLHS 309
 DB 275 VSOHIVFPYSTRSKSDHEELSLVASEAVRAIDKTSKTRTHGSGETLYLAPBGSGD 334
 QY 310 YSQLMTYPGYGVKAPDAEELDKVAKALAKALASVS-GTEYVGPECTCTYVPASGSSID 368
 DB 335 WIYDGIKYSFT 346
 QY 369 WAYDNGIKRFAFT 380

RESULT 15
 CBP2_SIMV1 STANDARD; PRT; 304 AA.
 ID CBP2_SIMV1 STANDARD; PRT; 304 AA.
 AC P42788;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
 OS Simulium vittatum (Black fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
 OC Chironomidae; Simuliidae; Simulium.
 OX NCBI_TaxID=7192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Gut;
 RA MEDLINE=94093864; PubMed=8269093;
 RA Ramos A., Mahowald A., Jacobs-Jorena M.;
 RT "Gut-specific genes from the black fly Simulium vittatum encoding
 RT trypsin-like and carboxypeptidase-like proteins.";
 RL Insect Mol. Biol. 1:149-163(1993).

CC -1- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
 CC -1- TISSUE SPECIFICITY: GUT-SPECIFIC.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L08481; AA018531.1; -.
 DR HSSP: P48052; IAYE.
 DR InterPro: IPR000834; Zn_carboxpept.
 DR Pfam: PF00246; Zn_carboxpept; 1.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolyase; Carboxypeptidase; Metalloprotease; zinc.
 FT NON_TER 1
 FT METAL 58 58 ZINC (BY SIMILARITY).
 FT METAL 61 61 ZINC (BY SIMILARITY).
 FT METAL 184 184 ZINC (BY SIMILARITY).
 FT ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 125 148 BY SIMILARITY.
 SQ SEQUENCE 304 AA: 34849 MW: 28663FF8A6A9144 CRC64;

Query Match 21.2%; Score 404.5; DB 1; Length 304;
 Best Local Similarity 33.0%; Pred. No. 7.4e-27;
 Matches 93; Conservative 52; Mismatches 86; Indels 51; Gaps 9;

QY 122 OYHSINEIYSWIEFTEHPDMLTKIHIGSSFEKYPPLVLYKSGKEQTAKNAIWDICGIIH 181
 DB 1 OYHTELEISWIDRLVOEHREHPEVYVGKSYEGEIRGVKYSYK--GNPVWVESNH 58
 QY 182 AREWISPA--FCL-----WFI-----GH-----NRMRKNS 206
 DB 59 AREWITAAATTYLLLELLTSKNSITREMAENDYVIFPVTPNDGYVYTHTRMRKTR 118
 QY 207 FYANNHCIGTDLSNPFVSKHMCESGSSSCSEYTCGLYPESEPEVAVASFLRRNINQI 266
 DB 119 PNPDSLCAGTDNRNRW-NFHMMEQIGSSRPCTETYGKKAPEVEETRSDTLKILKQI 177
 QY 267 KAYISMHSYOHIVFPYSTRSKSDHEELSLVASEAVRAIDKTSKTRTHGSGSETLY 326
 DB 178 KYVLAHFSYSQLLPEPYGHTCQHTYNHDDLOAIGDAARSLAQ-RYGDYTVGINIDALY 236
 QY 327 LAPGGDDWY---DLGIKYSFTSNP-----PVEKLTP 356
 DB 237 PASGSGMDWAVDTLPDIPIAYIELRPRDGMNGFOLPANOIIP 278

Search completed: September 18, 2002, 16:47:48
 Job time: 280 sec

OM of: US-09-980-881-2 to: GenEmbl: * out_format : pfs
Date: Sep 18, 2002 6:53 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-O=/cgn2_1/USPRO.spool/US09980881/runat_16092002.140128.10635/app-query.fasta.1.1126
-DB=GenEmbl -QPMF=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=1.000 -LOOPEXT=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELTOP=6.000
-DELXT=7.000 -START=1 -MATRIX=Diosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODAL=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIDE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09980881@cgn1.1.5320
-NCP=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPRY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-2
Query length: 360
Database: GenEmbl: *
Database sequences: 1797656
Database length: 187333701
Search time (sec): 4324.800000

Score_list:	Strid	Orig	ZScore	EScore	Len	Documentation
Sequence						
gb.pr:AB011969	+ 1884.00	3351.41	2.6e-178	1546	AB011969 Homo sapiens mRNA for	
gb.pr:AB086324	+ 1832.00	3260.29	3.1e-173	1272	AB086324 Sequence 1 from patent	
gb.pr:HMWCPX	+ 1832.00	3357.13	4.6e-173	1728	M75106 Human prepro-plasma caten	
gb.pr:116100	+ 1832.00	3357.01	4.7e-173	1749	116100 Sequence 2 from patent	
gb.pr:133526	+ 1832.00	3357.01	4.7e-173	1749	133526 Sequence 2 from patent	
gb.pr:BC007057	+ 1827.00	3248.25	1.4e-172	1715	BC007057 Homo sapiens, carboxy	
gb.pr:AB021968	+ 1534.50	2726.80	1.7e-143	1421	AB021968 Mus musculus mRNA for	
gb.pr:AF164524	+ 1534.50	2726.81	1.8e-143	1490	AF164524 Mus musculus thrombin	
gb.pr:AF186188	+ 1528.50	2725.49	6.8e-143	1430	AF186188 Mus musculus carboxy	
gb.pr:AB042598	+ 1502.50	2668.96	2.7e-140	1425	AB042598 Rattus norvegicus mRN	
gb.pr:AF190274	+ 700.50	1233.25	2.5e-60	1327	AF190274 Bothrops jararaca caten	
gb.pr:SSC133775	+ 646.00	1126.82	6.3e-55	1251	AJ133775 Sus scrofa falcipar	
gb.pr:HS224866	+ 641.00	1126.82	2.1e-54	1310	AJ224866 Homo sapiens mRNA for	
gb.pr:BC015338	+ 640.50	1126.70	2.4e-54	1454	BC015338 Homo sapiens, clone M	
gb.pr:167698	+ 639.00	1123.61	3.2e-54	1263	167698 Sequence 1 from patent	
gb.pr:AS1896	+ 639.00	1123.61	3.2e-54	1263	AS1896 Sequence 60 from patent	
gb.pr:AB085861	+ 639.00	1123.61	3.2e-54	1263	AB085861 Sequence 11 from patent	
gb.pr:AS1913	+ 631.50	1110.01	1.8e-53	1284	AS1913 Sequence 77 from patent	
gb.pr:AB085875	+ 631.50	1110.01	1.8e-53	1284	AB085875 Sequence 77 from patent	
gb.pr:AB085875	+ 628.50	1103.44	4.2e-53	1442	J05118 Mouse mast cell carboxy	
gb.pr:AB085875	+ 627.00	1102.22	5.0e-53	1251	J05118 Mouse mast cell carboxy	
gb.pr:BC012613	+ 625.50	1096.53	1.0e-52	1674	BC012613 Homo sapiens, Stimulat	
gb.pr:AB07368	+ 623.00	1089.45	2.6e-52	2154	AB07368 Sequence 124 from Paten	
gb.pr:AX333659	+ 621.50	1091.72	1.9e-52	1332	AX333659 Sequence 4168 from Paten	
gb.pr:HMWCPX	+ 621.50	1091.72	1.9e-52	1332	M81057 Human procarboxypeptid	
gb.pr:HMWCPX	+ 621.00	1090.11	2.3e-52	1427	U67914 Rattus norvegicus mast	
gb.pr:HMWCPX	+ 619.50	1086.11	3.9e-52	1622	M27717 Human mast cell carboxy	
gb.pr:G28614	+ 619.50	1086.11	3.9e-52	1622	G28614 human STR SHGC-35784, s	
gb.pr:AX106749	+ 611.50	1073.97	1.9e-51	1311	AX106749 Sequence 9 from Paten	
gb.pr:AR072911	+ 604.00	1064.11	6.6e-51	927	AR072911 Sequence 5 from Paten	
gb.pr:GCGBPARR	+ 581.00	1019.13	2.1e-48	1339	X64539 G.domesticus mRNA for c	
gb.pr:167699	+ 576.00	1014.03	4.1e-48	921	167699 Sequence 3 from patent	
gb.pr:AX106753	+ 564.00	983.90	1.9e-46	2128	AX106753 Sequence 13 from patent	
gb.pr:AX083139	+ 563.00	983.69	2.0e-46	1826	AX083139 Sequence 42 from Paten	
gb.pr:AS1908	+ 561.00	986.32	1.4e-46	999	AS1908 Sequence 72 from Paten	
gb.pr:AR085871	+ 561.00	986.32	1.4e-46	999	AR085871 Sequence 67 from Paten	
gb.pr:AS1904	+ 561.00	985.78	1.5e-46	1053	AS1904 Sequence 68 from Paten	
gb.pr:AR085868	+ 554.00	973.78	7.6e-46	1053	AR085868 Sequence 64 from Paten	
gb.pr:AX000417	+ 554.00	973.24	7.6e-46	1053	AX000417 Sequence 27 from Paten	

gb.pr:AX000418 - 554.00 973.24 7.6e-46 1053 | AX000418 Sequence 28 from p
gb.pr:AS1917 + 554.00 973.18 7.6e-46 1059 | AS1917 Sequence 81 from Pat
gb.pr:AR085878 + 554.00 973.18 7.6e-46 1059 | AR085878 Sequence 74 from p
gb.pr:AS1921 + 553.00 971.39 9.6e-46 1059 | AS1921 Sequence 85 from Pat
gb.pr:AR085881 + 553.00 971.39 9.6e-46 1059 | AR085881 Sequence 77 from p

seq_name: gb.pr:AB011969

seq_documentation_block:

LOCUS AB011969 1546 bp mRNA linear PRI 02-FEB-2000

DEFINITION Homo sapiens mRNA for carboxypeptidase B-like protein, complete

cds.
AB011969.1 GI:6855463
carboxypeptidase B-like protein.

KEYWORDS
Homo sapiens cDNA to mRNA.

SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (sites)
Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
Isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma

J Biol Chem. 266 (32), 21833-21838 (1991)

2 (sites)
Matsumoto,A.
Isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma

Unpublished (1998)

3 (bases 1 to 1546)
Matsumoto,A.
Direct Submission
Submitted (05-MAR-1998) Akira Matsumoto, Kobe University School of
Medicine, Department of Radiation Biophysics & Genetics;
Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan
(E-mail:amatedmed.kobe-u.ac.jp, Tel:81-78341-7451)

Sequence updated (10-Apr-1998).

location/Qualifiers
1. 1546
/organism="Homo sapiens"
/db_xref="taxon:9606"

20. 1102
/note="carboxypeptidase B-like protein in human brain"

/codon_start=1
/product="carboxypeptidase B-like protein"

/protein_id="BA090475.1"

/db_xref="GI:6855464"

/translation="MKLCISLAVVPIVLRCEQVHVPFQEGVIALIPRSROYVLON
LWTVYVILMQPTADILVKKGVHFFVNASDVNKAHLNVSIGCSYLAVEDLI
QQQISNDVSPRASASYEDYHLEINIEIETRHEDMLTKIHIGSFTEKYPDY
LKYSGKEQAKNAIWDICGHAHEWISPAECFLIGHNRMRKRNSFYANNHIGIDL
NRKFSKHCNECEGASSSCSEYCGLYPESEPVKAVASPLRNRINQIKAYISMHSYS
KHVPEYSTRSKSDHELSIIVASVAIAIEIKSNTRTRTHGSGETILAFGGDD
WYIDLIKTSFTSNPVEKLLPLSLK"

polya_site
1546
/note="21 A nucleotides"

BASE COUNT 453 a 323 c 297 g 473 t

ORIGIN

alignment_scores: Quality: 1884.00 Length: 360

Ratio: 5.263 Gaps: 0

Percent Similarity: 99.444 Percent Identity: 98.333

alignment_block:

US-09-980-881-2 x AB011969 ..

Align seg 1/1 to: AB011969 from: 1 to: 1546

|||||Met|y|S|e|u|C|y|S|e|u|e|A|l|a|V|a|l|l|e|u|a|P|r|o|l|l|e|V|a|l|l|e|u|P|H|e|C|y|G|I|17

|||||Met|y|S|e|u|C|y|S|e|u|e|A|l|a|V|a|l|l|e|u|a|P|r|o|l|l|e|V|a|l|l|e|u|P|H|e|C|y|G|I|17

|||||Met|y|S|e|u|C|y|S|e|u|e|A|l|a|V|a|l|l|e|u|a|P|r|o|l|l|e|V|a|l|l|e|u|P|H|e|C|y|G|I|17

|||||Met|y|S|e|u|C|y|S|e|u|e|A|l|a|V|a|l|l|e|u|a|P|r|o|l|l|e|V|a|l|l|e|u|P|H|e|C|y|G|I|17

|||||Met|y|S|e|u|C|y|S|e|u|e|A|l|a|V|a|l|l|e|u|a|P|r|o|l|l|e|V|a|l|l|e|u|P|H|e|C|y|G|I|17

|||||Met|y|S|e|u|C|y|S|e|u|e|A|l|a|V|a|l|l|e|u|a|P|r|o|l|l|e|V|a|l|l|e|u|P|H|e|C|y|G|I|17

```

20 ATGAGCTTTCAGACCTTCGACCTTCCTGACCCATTTCTCTCTCTGTA 69
17 uGlnHlsValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
   |||||
70 GCAGCATGCTCTCGCGCTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTTA 119
34 rGthSerArgGlnValGlnValLeuGlnAsnLeuThrThrTYrGlu 50
   |||||
120 GAACCTCTAGGCAAGTTCAGATTCTACAGAACTTACTACACATATAGAG 169
51 lIleValLeuTPGlnProValThrAlaAspLeuIleValLysLysG 67
   |||||
170 ATTGTTCTCTGGCAGCGGTTACAGCTGACCTTATTGTGAAGAAAAACA 219
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaH 84
   |||||
220 AGTCATGTTTGTAAATGATCATCTGATGTCGACCAATGTGAAGCCCAT 269
84 eUAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGlu 100
   |||||
270 TAAATGTAGCGGAATTCATGCACTGCTCTGCGCAGATGTGGAAGAT 319
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
   |||||
320 CTATATCAACAGCAGATTTCCACAGACAGCTCAGCCCGGACCTCCGC 369
117 aSerTYrTYrGluGlnTYrHISerLeuAsnGluIleTYrSerTYrIleG 134
   |||||
370 ATCGTACTATGAACAGATACACTCACTAAATGAATCTATTTCTGTGAT 419
134 lUpheIleThrGluArgHisProAspMetLeuThrLysIleHISerIleG 150
   |||||
420 AATTATTAACGTAGAGCATCTGATATGCTTACAAAATCCACATTTGGA 469
151 SerSerPheGluLysTYrProLeuTYrValLeuLysValSerGlyLys 167
   |||||
470 TCCCTCATTTGAGAAATGCCACACTGATGTTTAAAGTTTCTGAAAAAGA 519
167 uGlnThrAlaLysAsnAlaIleThrIleAspCysGlyIleHISerAla 184
   |||||
520 ACNAGCAGCCAAAATGCCATATGATGATGATGATGATGATGATGATG 569
184 lUTrPLeSerProAlaPheCysLeuThrPheIleGlyHisAsnArgMet 200
   |||||
570 AATGGATCTCTCTGCTCTCTGCTTGTGCTTATAGCCATATTCAGAAAG 619
201 TrpArgLysAsnArgSerPheTYrAlaAsnAsnHIScysIleGlyThr 217
   |||||
620 TCGAGAAAGAACCGTTCTTCTATGCGAACAAATCATTCATCGAGAAC 669
217 PleuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSer 234
   |||||
670 CCGTAATAGGAACCTTGGCTTCCAAACACATGCTGAGGAAGGTGCATCA 719
234 eTserSerCysSerGluThrTYrCysGlyLeuTYrProGluSerGluPro 250
   |||||
720 GTTCCCATGCTCGGAACCTACTGTGACTTATTCCTGAGTCAGAACCA 769
251 GluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIle 267
   |||||
770 GAAGTGAAGGCGAGTGTGTTCTTGAAGAAATATCAACACAGATTAA 819
267 sAlaTYrIleSerMetHISerTYrSerGlnHISerIleValIlePhePro 284
   |||||
820 AGCATATCATCAGCATGATCATCTACCTCCAGCATATAGTGTTCATATT 869
284 eTYrTYrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
   |||||
870 CCTATATACAGAGTAAAGCAAGACATGAGAACTGCTCTAGTAGGCC 919
301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTYrTh 317
   |||||
920 AGTGAAGCAGTTCTGCTATTTGAGAAATTAAGTAAATAATACAGAGTAT 969

```

```

317 rHISGlyHISerGluThrLeuTYrLeuAlaProGlyGlyLysAsp 334
   |||||
970 ACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGGAGGTGGGAGC 1019
334 sPTriIleTYrAspLeuGlyIleLysTYrSerPheThrSerAsnPro 350
   |||||
1020 ATTGGATCTATGATTTGGGCATCAAAATATTCGTTTACATCAAAACCA 1069
351 ValGluLysLeuLeuProLeuSerLeuLys 360
   |||||
1070 GTAGAGAACTTTTGGCCGTCTCTTAAAA 1099
seq_name: gp_pat:AR086324
seq_documentation_block:
LOCUS AR086324 1272 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 1 from patent US 5985562.
ACCESSION AR086324
VERSION AR086324.1 GI:10013090
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1272)
AUTHORS
Morse, M. John and Nagashima, M.
TITLE
Method of detecting thrombotic disease risk associated with plasma
carboxypeptidase B polymorphisms
JOURNAL
Patent: US 5985562-A 1 16-NOV-1999;
FEATURES
location/qualifiers
source 1..1272
BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN
alignment_scores:
Quality: 1832.00 Length: 415
Ratio: 5.103 Gaps: 3
Percent Similarity: 86.506 Percent Identity: 85.783
alignment_block:
US-09-980-881-2 x AR086324 ..
Align seg 1/1 to: AR086324 from: 1 to: 1272
1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysG 17
   |||||
1 ATGAGCTTTCAGACCTTCGACCTTCCTGACCCATTTCTCTCTCTGTA 50
17 uGlnHlsValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
   |||||
51 GCAGCATGCTCTCGCGCTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTTA 100
34 rGthSerArgGlnValGlnValLeuGlnAsnLeuThrThrTYrGlu 50
   |||||
101 GAACCTCTAGGCAAGTTCAGATTCTACAGAACTTACTACACATATAGAG 150
51 lIleValLeuTPGlnProValThrAlaAspLeuIleValLysLysG 67
   |||||
151 ATTGTTCTCTGGCAGCGGTTACAGCTGACCTTATTGTGAAGAAAAACA 200
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaH 84
   |||||
201 AGTCATGTTTGTAAATGATCATCTGATGTCGACCAATGTGAAGCCCAT 250
84 eUAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGlu 100
   |||||
251 TAAATGTAGCGGAATTCATGCACTGCTCTGCGCAGATGTGGAAGAT 300
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
   |||||
301 CTATATCAACAGCAGATTTCCACAGACAGCTCAGCCCGGACCTCCGC 350

```

```

117 aSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleG 134
|||||
351 ATGCTACTATGACAGATACATCAGTCAATTAATCTATCTTGGATGAG 400
|||||
134 IuphellerhrgluAqghisProaspMethleuthrlyslleHisIleGly 130
|||||
401 AATTATTAATGAGAGGATCCGTATATGCTTACAAAATCCACATGGA 450
|||||
151 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysG 167
|||||
451 TCCTCATTTAGAGATACCCACTCTATGTTTAAAGGTTTCTGGAAAAGA 500
|||||
167 uGlnTrpAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
|||||
501 ACAACAGCCAAAATGCGCATATGATGATGACTGCGAATCCATGCCAGAG 550
|||||
184 IuTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
|||||
551 AATGGATCTCTCTGCTTCTGCTTGTGTTCAATAGCCCATATAACTCAA 600
|||||
197 ..... 197
601 TTCTATAGGATATAGGCAATATACCAATCTCTGAGGCTTGGATTT 650
|||||
197 ..... 197
651 CTATGTTATGCCGCTGTTAATGNGAGCGTTATGACTACTCATGAAAA 700
|||||
198 .. AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCys 213
|||||
701 AGATGAAAGTGTGAGAGAAAGAACCGTCTTCTATGGAACAATCATATGC 750
|||||
214 ILGGLTThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluG 230
|||||
751 ATGGAGACAGACTGTAATAGGAACCTTGGCTTCCAAACACTGGGTGGA 800
|||||
230 uGlyAlaSerSerSerSerCysSerGluTrpTyrCysGlyLeuTyrProG 247
|||||
801 AGGTGATCCAGTTCCTCATGCTCGAAGACCTACTGAGACTTATGCTGT 850
|||||
247 IuSerGluProGluValLysAlaValAlaSerPheLeuArgAlaIle 263
|||||
851 AGTCAGAACCGAAGTGAAGGAGGTGCTTCTTGAAGAAAGAAATATC 900
|||||
264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVa 280
|||||
901 AACCAAGTTTAAAGCATACATCAGCATGCTATCTACCCAGCATATAGT 950
|||||
280 IpheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuS 297
|||||
951 GTTTCATATTTCTTATACAGAGTAAGTAACCAAGACCATGAGAACTGT 1000
|||||
297 erLeuValAlaSerGluAlaValAlaArgAlaIleAspLysThrSerLysAsn 313
|||||
1001 CTCTAGTAGGACGATGAGAGCTGCTCTATGTAAGAACTAGTAAAAAT 1050
|||||
314 ThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProG 330
|||||
1051 ACCAGGATACACATGGCATGGCTCAGAAACCTTATACCTAGCTCTCTGG 1100
|||||
330 yGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer...Phe... 345
|||||
1101 AGGTGGGAGCAGATGATGATGATTTGGGCATCAATATATTCGTTTACAA 1150
|||||
346 .....Th 346
1151 TTGAACCTTCAGATACGGGACATACGAGATTCTTGTCCGCGAGCGTTAC 1200
|||||
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
|||||
1201 ATCAAACCCACCTGTAGAGAAAGCTTTTGGCGCTGTCTCTAAAA 1243
|||||

```

```

seq_name: gb_pr:HMPCPBX
seq_documentation_block:
LOCUS       HMPCPBX                               1728 bp    mRNA    linear    prt 07-JAN-1995
DEFINITION  Human prepro-plasma carboxypeptidase B mRNA, complete cds.
ACCESSION   M75106
VERSION     M75106.1 GI:189686
KEYWORDS    plasma carboxypeptidase.
SOURCE      Homo sapiens liver cDNA to mRNA.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 1728)
AUTHORS     Eaton,D.L., Mallory,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
TITLE       Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
JOURNAL     J. Biol. Chem. 266 (32), 21833-21838 (1991)
MEDLINE     92042093
FEATURES
source      Location/Qualifiers
            1..1728
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /tissue_type="liver"
            1..1728
            /gene="PCPB"
            20..1291
            /gene="PCPB"
            /note="prepro-plasma carboxypeptidase B"
            /codon_start=1
            /protein_id="AAA60042.1"
            /db_xref="GI:189687"
            /translation="MKLSLAVLPVIVLVEQHFVAFSGQVLAALPRTSRQVQLQN
            LTTVEIYLMOPTADLIYKKQVHFVNAQVNDVKALNLSGILPCSVLADVEDLI
            QOOISNDYSPRASPASYEORYSLNEIYSIMFETREHPDMFTKHISSEPKPLVY
            LKSKSEORAKNAIVIDCSIAHREHISAPFCMPRTGHTQPGYGIQYTNLRLADPY
            VAPVYVNDGTDIDSMKRNRMKRNRSFTANNHCTIGDLNRFASKHCEGASSSCSE
            TYCGLYPESEPEVKAVASFLRNINQIAVISMHSYSOHIVPYSPYSTRKSKDHELS
            IVASAVRAIERTSKNRTYTHGSGSETIYLAIGGGDDWIDYLGIRYSFTIELRDYGT
            YGLPERYIKPTCRBAFAVSKIAHVIHNV"
            86..1288
            /gene="PCPB"
            /product="plasma carboxypeptidase B"
BASE COUNT  518 a      354 c      338 g      518 t
ORIGIN
alignment_scores:
Quality: 1832.00      Length: 415
Ratio: 5.103          Gaps: 3
Percent Similarity: 86.506      Percent Identity: 85.783
alignment_block:
US-09-980-881-2 x HMPCPBX
Align seg 1/1 to: HMPCPBX from: 1 to: 1728
1 MetLysLeuCysSerLeuAlaValIleValProIleValLeuPheCysG 17
|||||
20 ATGAAGCTTTCACACCTTCGACGCTCTGTACCCATGTGCTCTCTGTGA 69
|||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
|||||
70 GCACCATGTCTTCGCGTTTCAGAGTGGCCAAAGTCTGAGCTCTCTCTTA 119
|||||
34 rGThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGlu 50
|||||
120 GAACCTCTAGGCAAGTTCAGATCTTACAGAACTTACTACAAATATAGAG 169
|||||
51 ILeValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGly 67
|||||
170 ATTGTTCTCTGGACCGCGGTACAGCTGACCTTATTGTGGAAGAAAAACA 219
|||||
67 nValHisPhePheValAsnAlaSerAspValAlaAspAsnValLysAlaHisL 84
|||||

```

```

|||||
220 AGTCATTTTGTGTAATGATCTGATGTCGACAAATGTAACCCCATTT 269
84 euAsnValSerGlyIleProCysSerValIleuLeuAlaAspValIleuAsp 100
270 TAAATGTGAGCGGAAATTCATGACGTCTTCTGCGGACAGCTGGAAGAT 319
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaGalaSerAl 117
320 CTATATTCACAGCAGATTTCCACAGACAGATGAGCCCGGAGCTCCGCG 369
117 aSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleG 134
370 ATGCTACTATGACACAGATATCATTCTAAATGAATCTATCTTGGATAG 419
134 IuPheIleThrGluArgHisProAspMetLeuThrIleHisIleGly 150
420 AATTTAACTGAGAGGACATCTGATATGCTTACAAAATCCACATTCGA 469
151 SerSerPheGluIuSerTyrProLeuTyrValIleuLysValSerGlyLysG 167
470 TCCTCATTTGAGAGTACCCACTCTATGTTTAAAGGTTCTGGAAGAA 519
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
520 ACAACAGCCCAAAATGCCATATGATGATGCTGGAATCCATGCCAGAG 569
184 IuTrpIleSerProAlaPheCysLeuTrpPheIleGlnHis 197
570 AATGATCTCTCTCTCTCTCTCTGTTGATGAGCCATATTAACTCAA 619
197 ..... 197
620 TTCTATGGGATATAGGCAATATACCAATCTCTGAGGCTGTGGATTT 669
197 ..... 197
670 CTATGTTATGCCGGTGAATGTCGACGGTATGACTACTATGAGAAA 719
198 .. AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCys 213
720 AGAATCTGAATGTGGAGAAAGAACCCCTTCTTCTATGCCGAACATCACTGC 769
214 ILeGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluG 230
770 ATCGGAACGACCTGATAGAACTTCTCTCCAAACACTGGTGAGGA 819
230 uGlnLysSerSerSerCysSerGluIuThrTyrCysGlyLeuTyrProG 247
820 AGCTGATCCAGTCTCTCATGCTCGGAACCTACTGTGACCTTTATCTCG 869
247 IuSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
870 ACTCGAACCCAGAGTGAAGGCGAGTGGCTAGTCTTCTGAGAGAAATATAC 919
264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
920 AACGAGATTAAGCATATCATCATGATTCATATCTCCACACATATTAAT 969
280 IuPheTrpTyrSerTyrTrpArgSerLysSerLysAspHisGluGluLeu 297
970 GTTTTCCAAATTCCTATACAGCAAGTAAAGCAAAAGACCATGAGAACTGT 1019
297 eLeuValAlaSerGluAlaValAlaArgAlaIleAspLysThrSerLysAsn 313
1020 CTCTAGTACCCAGTGAAGCAGTTCGCTATTTGAGAAAACTAGTAAAT 1069
314 ThrArgTyrTrpHisGlyHisGlySerGluTrpHisLeuTyrLeuAlaProG 330
1070 ACGAGGTATACATGATGCCATGGCTCAGAAACCTTATATACGCTCCG 1119
330 yGluGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe... 345
|||||

```

```

1120 AGGTGGGAGCATTTGATCTATGATTTGGGCATCAATATTCGTTACAA 1169
346 .....Tn 346
1170 TTGAACCTTCAGATACGGGCACATACGATTCCTGCGGAGCGCTTAC 1219
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
1220 ATCAAAACCCACCTGTACAGAAAGCTTTGGCGGTGTCTTAAAA 1262

seq_name: gb_pat:116100
seq_documentation_block:
LOCUS 116100 1749 bp DNA linear PAT 03-APR-1996
DEFINITION Sequence 2 from patent US 5474901.
ACCESSION 116100
VERSION 116100.1 GI:1251008
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1749)
AUTHORS Dryden,D.T. and Eaton,D.L.
TITLE Antibodies to human carboxypeptidase B and methods of use thereof
JOURNAL Patent: US 5474901-A 2 12-DEC-1995;
FEATURES
source 1..1749
BASE COUNT 521 a 361 c 342 g 525 t
ORIGIN

Alignment_scores:
Quality: 1832.00 Length: 415
Ratio: 5.103 Gaps: 3
Percent Similarity: 86.506 Percent Identity: 85.783

Alignment_block:
US-09-980-881-2 x 116100 ..

Align seg 1/1 to: 116100 from: 1 to: 1749

1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysG 17
|||||
41 ATGAAGCTTTCGACGCTTCGACAGTGGCCAGTCTTACCTGCTCTCTGA 90
17 uGlnHisValPheAlaPheGlnSerGlyGlnValIleuAlaLeuProA 34
|||||
91 GCAGCATGCTTCGCGCTTTCAGAGTGGCCAGTCTTACCTGCTCTCTGA 140
34 rGThSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGlu 50
|||||
141 GAACTCTAGGCAAGTTCAGATTCTACAGAACTTACACAAACATATGAG 190
51 ILeValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG 67
|||||
191 ATTGTTCTCTGCGACCGGTAACAGCTTATGTGAGAAAAACA 240
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHis 84
|||||
241 AGTCATTTTGTAAATGCATCTGATGTGACACATGTGAAGCCCTT 290
84 euAsnValSerGlyIleProCysSerValIleuLeuAlaAspValGluAsp 100
291 TAAATGTGAGCGGAATTCATGACAGTGTCTGCTGGCAGAGTGAAGAT 340
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaGalaSerAl 117
|||||
341 CTATATCAACAGCAGATTTCCACAGACAGTCAAGCCCGGAGCTCCG 390
117 aSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleG 134
|||||
391 ATCGACTATGACAGTATCATCTACCTAAATGAATATTTCTTGATAG 440
|||||

```

```

134 lurbellierhgluarhisprompmetleuthrlslehllely 150
|||||
441 AATTATTAAGTGAAGGACATCTGATATGCTTACAAAAATCCACATTGGA 490
|||||
151 SerSerPheglulysylrproleuylrvalleuylsersglysgl 167
|||||
491 TCCTCATTTGAGAGTACCCACTATAGTTTAAAGGTTCTTGGAAGA 540
|||||
167 uglInthralalysasnalalietrpileaspcsglyllehlslalarg 184
|||||
541 ACAAGACGCAAAAATGCCATATGATGACTGTGGATTCATGCCAGAG 590
|||||
184 lurtprileserrproalaphecysleutrphelellyhlis 197
|||||
591 AATGATCTCTCCGCTTCTGCTTGCTTGCTGATGAGCCATATATCA 640
|||||
197 ..... 197
641 TTCTATGGATATATAGGCAATATACCAATCTCTGAGGCTTGATGATT 690
|||||
197 ..... 197
691 CTATGTTATGCCGTGTTATATGACGGTATGACTACTCATGCAAAA 740
|||||
198 ..AsnArgmetTrpArgLysAsnArgSerPheylrAlaAsnAsnHisCys 213
|||||
741 AGAATCGAAATGTGGAGAAAGAACCGTCTTCTATGCAACAAATCATGTC 790
|||||
214 lileglyhraspleuasnserasphevalserylsistrpcysglugl 230
|||||
791 ATCGGAACAGACTGAATAGAGACTTGTCCCAACACATGGTGTGAGA 840
|||||
230 uglYlaserSerSerSerCysSergluThrTyCysglyleutyrprog 247
|||||
841 AGGTGATTCAGTCTCTCATGCTGGAACCTACTGTGGACTTTATCTGT 890
|||||
247 lusergluprogluvallysAlaValalaserPheleuArgArgAsnIle 263
|||||
891 AGTCAGAACAGAGAGTGAAGGCACTGCTACTTCTTGAGAAAGAAATATC 940
|||||
264 AsnInlellysAlaTyrlleSerMetHisSerTySerGlnHisIleVal 280
|||||
941 AACCAAGTTTAAAGCATACATCAGCATTCATCTCCAGCATATAGT 990
|||||
280 lPheProtyrSerTyThrArgSerLysSerLysAspHisglulenus 297
|||||
991 GTTTCATATTCCTATACAGAACTAAAGCAAGACCATGAGGAACGT 1040
|||||
297 erleuValalasergluAlaValalargAlalaeasplysThrSerLysAsn 313
|||||
1041 CTCTAGTAGCCAGTGAAGCACTGCTATGTGAAAAACTAGTAAAAAT 1090
|||||
314 ThrArgTyThrHisGlyHisGlySerGluThrleuTyrlleuAlaProgl 330
|||||
1091 ACCAGGATATACATGAGCCATGCTCAGAAACCTTATACCTAGCTCCGG 1140
|||||
330 yglYlAspAspTrpIleTyAspleuGlylleystySerPhe... 345
|||||
1141 AGGTGGGAGAGATGATCTATGATTTGGGATCAAAATATTCGTTTCAA 1190
|||||
346 .....Th 346
1191 TTGAACCTTCGAGATACGGGCACATACGAAATCTTGCTGCCGAGCGTTAC 1240
|||||
346 rSerAsnProProValglulysleuLeuProleuSerleuLys 360
|||||
1241 ATCAACCCACCTGTAGAGAGCTTTGCCCCGCTGTCTATAAA 1283
|||||
seq_name: gb_pat:133526
seq_documentation_block:
LOCUS I33526 1749 bp DNA linear PAT 06-FEB-1997

```

```

DEFINITION Sequence 2 from patent US 5593674.
ACCESSION I33526
VERSION I33526.1 GI:1824317
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1749)
AUTHORS Drayna,D.T. and Eaton,D.L.
TITLE Plasma carboxypeptidase
JOURNAL Patent: US 5593674-A 2 14-JAN-1997;
FEATURES
Source
BASE COUNT 521 a 361 c 342 g 525 t
ORIGIN
alignment_scores:
Quality: 1832.00 Length: 415
Ratio: 5.103 Gaps: 3
Percent Similarity: 86.506 Percent Identity: 85.783
alignment_block:
US-09-980-881-2 x I33526 ..
Align seg 1/1 to: I33526 from: 1 to: 1749
1 MetLysleuCysSerleuAlaValleuValprolleuValleuPheCysgl 17
|||||
41 ATGAAGCTTTCAGGCTTCGACGTCCCTGTACCCATGTTCTCTGTGGA 90
|||||
17 uglHisValPheAlaPheGlnSerGlyGlnValleuAlaAlaLeuPro 34
|||||
91 GCAGATGCTCTTCGGCTTCAGAGTGGCCAGTGTACTGTCTCTCTGTA 140
|||||
34 rgnrSerArgGlnValGlnValleuGlnAsnleuThrThrTyrglu 50
|||||
141 GAACTCTAGGCAAGTCAAGTTCACAGAAATCTTACTACCAACATATGAG 190
|||||
51 lleValleuTrpGlnProValThrAlaAspleuIleValLysLysgl 67
|||||
191 ATTGTTCTGTGGACCCGTTAACCTGACCTTATTTGAGAGAAAAACA 240
|||||
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
|||||
241 AGTCATATTTTTTGAATGATGATGATGTCAGCAATGTGAAGCCATT 290
|||||
84 euAsnValSerGlylleProCysSerValleuLeuAlaAspValgluasp 100
|||||
291 TAAATGTAGCGGGAATTCATGCACTGCTGCTGGCAGACGTGGAGAT 340
|||||
101 leuilegInGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
341 CTATTCACACAGCAGATTTCACAGACAGACAGACCCCCGAGCCTCCGC 390
|||||
117 aSerTyTyrgluglnTyHisSerleuAsngluIleTySerTrpIleg 134
|||||
391 ATCGTACATGAGAGTATCACTCACTAAATGAATCATTTCTTGATAG 440
|||||
134 lurbellierhgluarhisprompmetleuthrlslehllely 150
|||||
441 AATTATTAAGTGAAGGACATCTGATATGCTTACAAAAATCCACATTGGA 490
|||||
151 SerSerPheglulysylrproleuylrvalleuylsersglysgl 167
|||||
491 TCCTCATTTGAGAGTACCCACTATAGTTTAAAGGTTCTTGGAAGA 540
|||||
167 uglInthralalysasnalalietrpileaspcsglyllehlslalarg 184
|||||
541 ACAAGACGCAAAAATGCCATATGATGACTGTGGATTCATGCCAGAG 590
|||||
184 lurtprileserrproalaphecysleutrphelellyhlis 197
|||||

```

```

|||||
591 ATGATCTCTCTGCTTCTGCTTGCATAGGCCATATACTCA 640
197 ..... 197
641 TTCTATGGATATAGGCAATATACCAATCTCTGAGGCTTGATTT 690
197 ..... 197
691 CTATGTTATGCCGGTGAATATGTCAGAGGTTATGACTCATGAGAAA 740
198 .. AaATgMeTtPArIglYAsnArSserPheTyrlAlaAsnArhIcYs 213
|||||
741 AGAATCGAATGTGGAGAAAGAACCGCTTCTTATGCGAACATCATGTC 790
214 TLeGlyThrAsPLeuAAsnSerAsnPheAlSerLysHsTrpCysGlU 230
|||||
791 ATCGAGACAGACCTGATAGGAATTTGCTCCAAACACGtGtGAGGA 840
230 uGlyAlaSerSerSerSerSerSerSerSerSerSerSerSerSer 247
841 AGGTCATCCAGTCCATCCATGCTCGAACCCTACTGCTTATCTCTG 890
247 TuseRgluprOgluValIyAlaValAlaSerPheLeuArGaArhIle 263
|||||
891 AGTCAGAACAGAAAGTAAGGAGCTGCTAGTTCCTTGAGAGAAATATC 940
264 AsnGlnIleLysAlaTyrlleSerMetHsSerTyrlSerGlnHsIleVa 280
941 AACAGATTAAGCATATACATGATGATCATCTACTCCAGCATATATG 990
280 lPheProTySerTyThrThrSerSerSerSerSerSerSerSerSer 297
991 GTTTCATATCTCTATACAGAAAGTAAAGCAAGACATGAGAACTG 1040
297 eRLeuValAlaSerGluAlaValAlaIleAspLysTrhSerLysAsn 313
|||||
1041 CTCTAGTAGCAGTGAAGAGCTCGTATATGAAAACTACTATAAAAT 1090
314 ThArGTYThrHsGlyHsGlySerGluThrLeuTyrlleuAlaProG 330
|||||
1091 ACCAGATTAACATGATGCTCGAAGACCTTATACCTACTCTCTG 1140
330 yGlyGlyAspAspTrpIleTyrlAspLeuGlyIleLysTrSer...Phe... 345
|||||
1141 AGCTGGCAGCATGATGATGATTTGGCATCAATATGCTTACAA 1190
346 .....Th 346
1191 TTGAACCTTCAGATACGGGCACATACGATTTCTGCTGCCGAGCTTAC 1240
346 rSerAsnProProValGlnLysLeuLeuProLeuSerLeuLys 360
|||||
1241 ATCAAAACCACTGTAGAGAGCTTTGCCGCTGCTCTTAAAA 1283

seq_name: gb.Pr:BC007057

seq_documentation_block:
LOCUS BC007057 1715 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
IMAGE:3934520, mRNA, complete cds.
ACCESSION BC007057
VERSION BC007057.1 GI:13937896
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (30-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer

```

REMARK

COMMENT

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONTECH
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mdépxall@stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 16 Row: f Column: 21
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4503004.

FEATURES

source

```

1..1715
/organism="Homo sapiens"
/db_xref="locusID:1361"
/db_xref="taxon:9606"
/clone="MGC:12495 IMAGE:3934520"
/tissue_type="Skeletal Muscle"
/clone_lib="NIH MGC_81"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
25..1296
/codon_start=1
/product="carboxypeptidase B2 (plasma)"
/protein_id="AA07057.1"
/db_xref="GI:13937897"
/translacion="MKLCSLAVLPVLYFCEQHYFAFGSQVLAALPRSRQVYLQ
LTTYTELVLPQVYADLIVRKQVHEPVNADVDVNAHNVAGIAPCSVLAVEDLI
QOOISNDVSPRASASYEQYHSEINSEISWIEFTERRPMLTKIHIGSSFEKPYLY
LKVSGEQAARKNAIMIDCGIARWISIPARCLMEFICHTIQFGITGOYTLALNPFY
VMPVAVNDGVDYSGWKRMKRMKRPYANNHCIGTDLNRPFAKHNCEBASSSCSSE
TYGGLYSESEPEKAVASPLRKNINQKATISMASHSIOHIVPSTYTRSKSKHSELS
LVASENRALEKSKNTRTHGHSELTLYLAPGGGDWIDGIDIKYSFTLELDGTLY
GFLPEPRTIPTRKFAVASKIAWHVIRNV"

```

CDS

```

BASE COUNT      518 a      348 c      341 g      508 t
ORIGIN
alignment_scores:
Quality: 1827.00      Length: 415
Ratio: 5.089      Gaps: 3
Percent Similarity: 86.506      Percent Identity: 85.542
alignment_block:
US-09-980-881-2 x BC007057

```

Align seg 1/1 to: BC007057 from: 1 to: 1715

```

1 MetLysIeucYsSerLeuAlaValLeuValProIleValIleuPheCysG 17
|||||
25 ATGAGCTTGTGAGCGCTTGAGCTCTGTACCAATGTTCTCTCTGGA 74
|||||
17 uGlnHsIValPheAlaPheGlnSerGlyGlnValIleuAlaIleuProA 34
|||||
75 GCAGCATGCTTCGGCTTTCAGAGTGCAGAGTCTACCTCTCTCTCTA 124
|||||
34 rGThSerArGlnValGlnValIleuGlnAsnIleuThrThrThrTyGlu 50
|||||
125 GAACCTTAGCGCAAGTTAAATTTACAGAAATCTTACACAACTATGAG 174
|||||
51 TLeValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG 67
|||||
175 ATTGTTCTCTGCGAGCGGTAAAGCTGACCTTATTTGGAAGAAAAACA 224

```

```

67  nvalHisPheValAlaSerValAlaAspValAlaHisL 84
68  |||||||
69  AGTTCATTTTGTAAATGATCTGATGAGCAATGTAAGCCCATTT 274
70  |||||||
71  eulsnValSerGlyIleProCysSerValLeuAlaAspValAla 100
72  |||||||
73  TAAATGAGCGAATTCATGCGAGTCTTTCGCGCAGACGTGGAAGAT 324
74  |||||||
75  CTTATTCACAGCAGATTTCCACAGACAGTACGCCGCCGAGCCCGC 374
76  |||||||
77  aSerTyTyGluGlnIleHisSerLeuAsnGluIleTySerTrpIle 134
78  |||||||
79  ATGCTACTAAGAACAGTATCTACATAATGAAATCTATCTTGATAG 424
80  |||||||
81  lupheIleThrGluArgHisProAspMetLeuThrIleHisIleGly 150
82  |||||||
83  AATTTATTAAGTGAAGGATCTGATATGCTTACAAAATCCACATTGGA 474
84  |||||||
85  SerSerPheGluIleTyTyProLeuTyValLeuIleValSerGly 167
86  |||||||
87  TCCCATTTGAGAGTACCCACCTATGTTTAAAGTTCTGGAAGA 524
88  |||||||
89  uGlnThrAlaIleAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
90  |||||||
91  ACAAGCGACCAAAATCCATATGATGATGCTGGAATCCATCCAGAG 574
92  |||||||
93  lurPheIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
94  |||||||
95  AATGATCTCTCCGCTTCTGCTTGCTTGCTGATAGGCCATATACATCA 624
96  |||||||
97  ..... 197
98  TTCTATGGAATATAGGCAATATACCAATCTCTGAGGCTTGATGATTT 674
99  |||||||
100 ..... 197
101 CTATGTTATGCGCGTGTAAATGATGATGATGATGATGATGATGATG 724
102 |||||||
103 .AsnArgMetTrpArgIleAsnArgSerPheTyAlaAsnHisCys 213
104 |||||||
105 AGAATCAATGTCGAGAAAGACCGTCTTCTTCTATGCGAACAATCATTC 774
106 |||||||
107 IleGlyThrAspLeuAsnSerAsnPheValSerIleHisIleTrpCysGlu 230
108 |||||||
109 ATCGAAGACAGCTGATAGGAATCTTCTTCAAAACAGCTGCTGAGGA 824
110 |||||||
111 uGlyAlaSerSerSerSerSerSerGluThrTyTyCysGlyLeuTyTrp 247
112 |||||||
113 AGGTGCAATCCAGTCTCTCATGCTCGAAACCTACTGCTGATTTATCT 874
114 |||||||
115 luserGluProGluValIleValAlaValAlaSerPheLeuArgArgAsn 263
116 |||||||
117 AGTCAGAACAGAAAGAGGACAGTGTGCTTCTTGAGAAAGAAATTC 924
118 |||||||
119 AsnGlnIleValAlaTyIleSerMetHisSerTySerGlnHisIleVal 280
120 |||||||
121 AACCAAGATTAAAGCATACATGATGATGATGATGATGATGATGATG 974
122 |||||||
123 lPheProTySerTyThrArgSerTySerTyAspHisGluGluLeu 297
124 |||||||
125 GTTCCATTAATCTTATACAGCAAGTAAAGCAAGACATGAGCAATGCT 1024
126 |||||||
127 erLeuValAlaSerGluAlaValAlaArgAlaIleAspTyThrSerTy 313
128 |||||||
129 CTCTAGAGCCAGTGAAGCAAGTGTGCTTATGAGAAACCTAGTAATAAT 1074
130 |||||||
131 ThrArgTyThrHisGlyHisGlySerGluThrLeuTyTrpAlaProG 330
132 |||||||
133 ACCAGGATATACATGCGCATGGCTCAGAAACCTATATACCATGCTCC 1124
134 |||||||
135 yGlyGlyAspAspTrpIleTyTrpAspLeuGlyIleValTySerPhe 345

```

```

1125 AGTGGGAGCATTTGATATGATTTGGCATCAAAATTCGTTTACAA 1174
1126 |||||||
1127 ..... Th 346
1128 ..... 11
1129 TTGAATCTCGAATACGGGACATACGATCTTCTCGCGAGCGTTAC 1224
1130 |||||||
1131 rSerAsnProValGluIleValLeuProLeuSerLeuIle 360
1132 |||||||
1133 ATCAAAACCCACTGTAGAGAGCTTTTCCGCTGCTCTAATAA 1267
1134 |||||||
1135 seq_name: gb_ro:AB021968
1136
1137 seq_documentation_block:
1138 LOCUS AB021968 1421 bp mRNA linear ROD 27-JUL-2000
1139 DEFINITION Mus musculus mRNA for carboxypeptidase R, complete cds.
1140 ACCESSION AB021968
1141 VERSION AB021968.1 GI:9558447
1142 KEYWORDS carboxypeptidase R.
1143 SOURCE Mus musculus
1144 ORGANISM Mus musculus
1145 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1146 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1147 TITLE 1 (sites)
1148 AUTHORS Sato,T., Miwa,T., Akatsu,H., Matsukawa,N., Ohta,K., Okada,N.,
1149 Campbell,W. and Okada,H.
1150 PRO-CARBOXYPEPTIDASE R IS AN ACUTE PHASE PROTEIN IN THE MOUSE,
1151 WHEREAS CARBOXYPEPTIDASE N IS NOT
1152 J. Immunol. 165 (2), 1053-1058 (2000)
1153 20341711
1154 JOURNAL
1155 MEDLINE Direct Submission
1156 REFERENCE Submitted (28-DEC-1998) Tomoo Sato, Fukushima Hospital, Choji
1157 AUTHORS Medical Institute: 19-14, Azayamanka, Noyori-cho, Toyohashi, Aichi
1158 TITLE 441-8124, Japan (E-mail:tomoo@cep-ip.or.jp, Tel:81-532-46-7511,
1159 Fax:81-532-46-4895)
1160
1161 FEATURES
1162 source location/Qualifiers
1163 1. 1421
1164 /organism="Mus musculus"
1165 /db_xref="taxon:10090"
1166 17..1285
1167 /product="carboxypeptidase R"
1168 /protein_id="BAB03402.1"
1169 /db_xref="GI:9558448"
1170 /translation="MKHLGILIVATIIYEOHGFAROSGOVLSAPRTSRQVLLQNL
1171 TTTFEVLNQPTAETIEKKEKVEHFVNASDVDSYKAHLNVSRIPTVNLNNVEDLIE
1172 OQFNDVSPBRASASYEQYHSINLETYSMEVTEQHPMLQKITYGSSFEKPYVL
1173 KVSGEORIKNAIWDICGIHAREMISPAFCIMVIGYTOFHGKENDYTRILRHVDYI
1174 MPYMANVDGYDTWKKNRMKRNKSAKNRNCVCTDLNRNPASKHCMEKGSSEST
1175 YGILYSEPEYVAADPLBRNDIHKAYISMHSYQOILPVSYSRKSKEDEEEL
1176 VASEAFAIESIKNRYTHSGSESLYLAPGSDMIYDILGIKYSFTIELRDTGRTG
1177 FLPERIKPTCEALALAIISKIYHWIYKNT"
1178 1421
1179 polyA_site /note="11 a nucleotides"
1180
1181 BASE COUNT 418 a 320 c 305 g 378 t
1182 ORIGIN
1183
1184 alignment_scores:
1185 Quality: 1534.50 Length: 415
1186 Ratio: 4.540 Gaps: 4
1187 Percent Similarity: 81.446 Percent Identity: 72.530
1188
1189 alignment_block:
1190 US-09-980-881-2 x AB021968 ..
1191
1192 Align seg 1/1 to: AB021968 from: 1 to: 1421
1193
1194 1 MetysleuCySerLeuAlaValLeuValProIleValLeuPheCysG 17
1195 |||||||
1196 ..... 11
1197 ATGAAGCTTCATGCCCTTGAAATCTCGTATGACCATCATCTCTAT...GA 63

```



```

/protein_id="AF00528.1"
/db_xref="GI:6003652"
/translation="MKLHGILIVAILTEYEQHGFAPGQVSLAPRTSRNOYOLLNL
TYYEVLMQPTAEFEKKEVHFEFVNASDVSVSAHLNVSIIPNVIMNVEDIE
OCTFNDFVRBASAYEBOYHSLNETYSNIEVTEBOHPMLQKITYIGSSPERPLIYL
KVGKROIRKNTWIDCGIHAHEMISPARCLNFTGYTQHGKENTLTLRLHVDYI
MPVNMDDGIDYTKKRMKRNKNSAHKNNKCVGTDLNRGASHWCEKASSSCSET
YGLIPSEPEVKAADPLRNIDHFKAYISMHSQOILFPYSYNRSKSKDHEELSL
VASEAVRAIGSEFNKNTRYTHGSGESLYLAPGSDWDIMYDLGIKISFTIELNDTGRYG
FLPERRYIKPTCAELAAISKIYMWVIRNT"
BASE COUNT      421 a      320 c      310 g      379 t
ORIGIN

alignment_scores:
    Quality: 1528.50      Length: 415
    Ratio: 4.536          Gaps: 4
Percent Similarity: 81.205      Percent Identity: 72.289

alignment block:
US-09-980-881-2 x AF186188 ..
Align seg 1/1 to: AF186188 from: 1 to: 1430

1 MetLysLeuCySerLeuAlaValLeuValProLeuValLeuPheCysG1 17
|||||.....|.....|.....|.....|.....|.....|
20 ATGAAGCTTCATGGCTTGATGCTGTGAGCCATCATCTCTAT...GA 66
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProa 34
|||||.....|.....|.....|.....|.....|.....|
67 GCACATGGCTGGCTGGCTTTCAGAGTGGCCAGTTTATCTGCTTCCAA 116
34 rGmTrSerAArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGlu 50
|||||.....|.....|.....|.....|.....|.....|
117 GAACCTCCAGGACAGTCACTACTCTCAGATCTTACTACACGATAGAG 166
51 lLeValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
|||||.....|.....|.....|.....|.....|.....|
167 GTGCTTCTCGGACGACGACGACGACGCTGAATTCCTCAGAAAGAAAGA 216
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaH1L 84
|||||.....|.....|.....|.....|.....|.....|
217 AGTCCACTTTTGTGTGATGCTGTGATGCTGACAGTGCAGGCGCAT 266
84 euAsnValSerGlyLleProCySerValLeuLeuAlaAspValGluAsp 100
|||||.....|.....|.....|.....|.....|.....|
267 TAAATGTGACGACAGAAATTCATTAACTCTGTGATGAACAACGAGGAG 316
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||.....|.....|.....|.....|.....|.....|
317 CTAATGTAACAGACAGACTTCATGACAGCGTACGCCCCGCCCTCCGC 366
117 aSerTyrTyrGlnGlnTyrHisSerLeuAsnGlnLyleTyrSerTrp166 134
|||||.....|.....|.....|.....|.....|.....|
367 TTCTATCTATGACGACATACATCAGCTGAATGAAATCTATCTGTGATAG 416
134 lueHeleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
|||||.....|.....|.....|.....|.....|.....|
417 AACTCTTAACGACAGCATCTGACATGCTCCAGAAATCTAACATCGGA 466
151 SerSerPheGlnLysTyrProLeuTyrValLeuLysValSerGlyLysG1 167
|||||.....|.....|.....|.....|.....|.....|
467 TCATCTATCGAGAGTACCCACTTATGTTTAAAGTCTCAGAGAAAGGA 516
167 uGlnThrAlaLysAsnAlaIleTrpLleAspCysGlyLleHisAlaArg 184
|||||.....|.....|.....|.....|.....|.....|
517 ACMAAGAAATCAAAATGCGCATCTGATCGACTGTGAATCATGACAGAG 566
184 lueTrpLleSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
567 AATGCAATTCACCTGCTTCTGTTGTGTGATGCTTACGCTACGACACAA 616
197 ..... 197

```

```

617 TTCATGGGAAAGAAATCTGTATACAGACTTCTGAGCCACGTCGATTT 666
197 ..... 197
667 CTACATCATGCCCGTGTATGAACGTGATGCTATGACTACACGTGAGAA 716
198 ..AsaArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCys 213
717 AGAATCCAAATGTGAGAGAGAACCGCTGCTGCACAGAAACACCGCTGC 766
214 lLeGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGlu 230
767 GTGGCAGACAGCTGAGACAGAACTTCTCCAAACACGCTGTGAGAA 816
230 uGlyAlaSerSerSerSerSerGlyLysTyrGlyLysLeuTyrProG 247
817 AGTGCCTCAAGTCCCTGCTGCTGAAACCTACTGCTGATTCCTG 866
247 lueSerGluProGluValLysAlaValAlaSerPheLeuArgAsnAlle 263
867 AGTCTGAGCCAGAGGTGAAGCGAGTGGCTGACTTCTTGAGAGAAATATC 916
264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVa 280
917 GACCACATTAAGCTTACATCAGATGATGCACTACTCCCAAAATATCT 966
280 lPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeu 297
967 GTTCCCTTAATCTTAAACGAAAGCAAGCAAGCAAGCAAGCAAGCAAG 1016
297 erLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsn 313
1017 CTCTAGAGGCGACGAGACGAGTTCGCAATTCGAAAGTTTAAATTAAC 1066
314 ThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProG1 330
1067 ACCAGGTACACACGACGAGTGGCTCAGAAAGTTTATCTAGCTCTCG 1116
330 yGlyGlyAspAspTrpLleTyrAspLeuGlyLysLysTyrSerPhe... 345
1117 AGCTTCTGACGATGATGATGATGATGATGATGATGATGATGATGAT 1166
346 ..... 1166
1167 TTGAGCTCCGAGATACAGCAGATACGATTCCTGCTGCTGAGAGATTC 1216
346 lSerAsnProProValGlnLysLeuLeuProLeuSerLeuLys 360
1217 ATCAAAACCCACTGTCAGAGACTTGGCCGCCATCTCTAATAA 1259

seq_name: gb_to:AB042598
seq documentation block:
LOCUS      AB042598      1425 bp      mRNA      linear      ROD 02-DEC-2000
DEFINITION Rattus norvegicus mRNA for pre-procarboxypeptidase R, complete cds.
ACCESSION  AB042598
VERSION    AB042598.1 GI:11526576
KEYWORDS
SOURCE
ORGANISM  Rattus norvegicus CDNA to mRNA.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (sites)
REFERENCE  Kato,T., Akatsu,H., Sato,T., Matsuo,S., Yamamoto,T., Campbell,W.,
AUTHORS   Hotta,N., Okada,N. and Okada,H.
TITLE     Molecular cloning and partial characterisation of rat
           procarboxypeptidase R and carboxypeptidase N
JOURNAL   Microbiol. Immunol. 44 (8), 719-728 (2000)
MEDLINE   20471387
REFERENCE  2 (bases 1 to 1425)
AUTHORS   Kato,T.

```

TITLE Direct Submission
JOURNAL Submitted (10-MAY-2000) Tomomi Kato, Choju Medical Institute, Choju Medical Institute, Noyori-cho, Yamanaka, 19-14, Toyohashi, Aichi 441-8124, Japan (E-mail: kato@tomotsuru.med.nagoya-u.ac.jp, Tel:0532-46-7511, Fax:0532-46-4899)
FEATURES Location/Qualifiers
source 1.1425
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 100.1368
 /codon_start=1
 /product="pre-procarboxypeptidase R"
 /db_xref="GI:11526577"
 /translation="MKLYGLVLAIIIEKKGKLAFOGSHVLSALPRTSROYOLLQNT
TTTVEVLMOPVTAERIEKKEVHEFVNADVNSKAYLNASRIPIPNVLMNVEDLLO
OOTNDVSPRASSSYEQHSLNEYSWIEVTEBQPMLOKITYGSSYEKPLVYL
KYSKEHRYKNAIWDGCIHAREWISPAFLWFGVTFQFHKEMTYTLRHVDYI
MPYAVDGDYTWKKRBMQRKRSYHNNRCYGTDLNRFASKHMEKASSTSCSEY
YCGLPSESEPEYKAVADFLRNINIKATISMSTSOQLFPSTIRSKSDHEEISL
VASEAVRALESINKNRYTHSGSSSLYLA PGSDWDIYDLGIKYSFTLELRDTGRYG
FLPPEFLKPTCAELAAVSKIAMVIRNS"

BASE COUNT 417 a 334 c 309 g 365 t
ORIGIN

alignment_scores:
 Quality: 1502.50 Length: 415
 Ratio: 4.445 Gaps: 4
Percent Similarity: 81.446 Percent Identity: 70.843

alignment_block:
US-09-980-881-2 x AB042598 ..

Align seg 1/1 to: AB042598 from: 1 to: 1425

```

1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCysG1 17
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 ATCAAGCTTATGAGCTTGGCTTGGCTGACCTCATCTCTAT...GA 146
17 uGlnHISValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
147 GAACATGAGGCTTCCCTTCAGAGTGGCCATGTTCTATCTCTCCCTC 196
34 rGthrSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrglu 50
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
197 GAACCTCCAGCAGATTCACACTCTTCAGAACTCAGACAACTTACAG 246
51 IleValLeuThrGlnProValThrAlaAspLeuIleValLysLysG1 67
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
247 GTTGTCTCTGGAGCCAGTGACAGCTGAATTCATTGGAAGAAAAAGA 296
67 nValHISphePheValAsnAlaSerAspValAspAsnValLysAlaHISL 84
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
297 AGTCCATCTCTTGTGAATGGGTGTGATGTAAAGTGTCAAAAGCTATT 346
84 euAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
347 TAAATGCGAGCAGAAATTCATTAACTCCGATGAACACGTGAGAGAT 396
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaGlnSerA1 117
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
397 CTAATTAACAGCAGACGTCCAAATGACACTGTGAGCCCGAGCTCTCTC 446
117 aserTyrgluGlnGlnIleThrIleSerLeuAsnGlnIleTyrgluSerP1leg 134
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
447 CTCCTACTATGACAGATATCACTGTTAAATGAATATCATTTCCGAGATG 496
134 IupheIleThrGlnArgHisProAspMetLeuThrLysIleHisIleGly 150
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
497 AAGTTAATAAGACACACCTGACATGCTCCAGAAAATCTACATTGCA 546
151 SererheGluLysTyrgluProLeuThrValLeuLysValSerGlyLysG1 167

```

```

547 TCCCTATATGAAAAGTACCACCTTATGTGTAAAGGCTCAGGAAAAGA 596
167 uGlnIleThrAlaLysAsnAlaIleThrP1leAspCysGlyIleHisAlaArg 184
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
597 ACACGAGATCAAAAATGCCATATGATGATGACTGTGGATTCATGCCAGAG 646
184 IuThrP1leSerProAlaPheCysLeuThrPheIleGlyHis..... 197
647 AGTGGATTTCACACAGCTTCTGCTGTGTTCAATAGCCTATGTAACGCA 696
197 ..... 197
697 TTCATGGGAAAAGAAATATACATACACAGACTTGTAGGACAGTGAGATT 746
197 ..... 197
747 CTCATTTATGCCAGATGATGATGTGGAGGCTAGACACTACAGTGGAAAA 796
198 ..AsnArgMetThrParGlyAsnArgSerPheTyrgluAlaAsnAsnHisCys 213
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
797 AGAATCGAATGTGGAGAAAGAACCCCTCTGTCCACATGAAACAACCGCTGC 846
214 IleGlyThrAspLeuAsnSerAsnAspValSerLysHisThrPcyGluG1 230
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
847 GTGGCAGACAGACTGAAACAGAACTTCCTCCAAACACTGGTGGAGAA 896
230 uGlyAlaSerSerSerCysSerGluThrTyrglyLeuTyrgluTyrglu 247
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
897 AGGCCATCAAGTTCCTCTCTGCTGTGAGACCTACAGTGCATTTACCTG 946
247 IuSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 263
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
947 AGTCGACCCAGAGGTGAAGCAGAGTGGCTGACTTCTGAGAGAAATATTC 996
264 AsnGlnIleLysAlaTyrlleSerMetHisSerTyrgluHisIleVal 280
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
997 AACCATTAAGCTTATCATCATGATGATGACATCATACTCCAGCAAAATCT 1046
280 IpheProTyrgluSerTyrgluThrArgSerLysSerLysAspHisGluLys 297
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1047 GTTCCCTATTCCTTCACACAGACGAAAAGCAAGACCAACAGAGAACTGT 1096
297 erLeuValAlaSerGluAlaValAlaArgAlaIleAspLysThrSerLysAsn 313
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1097 CTCGTAGTGGCCAGCAGAACAGATGCTGCAATGAAGATTAATTAATAAAC 1146
314 ThrArgTyrgluHisGlyHisGlySerGluThrLeuTyrgluAlaProG1 330
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1147 ACCAGGTACACACATGAGCAGTGGCTCAGAAAGTTATATCTAGCTCCG 1196
330 YGlyGlyAspAspThrP1leTyrgluAspLeuGlyIleLysTyrgluSerPhe 346
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1197 AGTTCATATGATGATGATGATTTGGGCATCAAAATTTCTGTTACGA 1246
346 ..... 346
1247 TTGAATCTCGGAGATACAGCAGATACGGGTTCTGCTCCTCAGAGATTTC 1296
347 ..SerAsnProProValGlnLysLeuLeuProLeuSerLeuLys 360
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1297 ATCAAAACCCACTTGGCAGAAAGCTTTGGCCGCGAGTCTTAAAA 1339
seq_name: gb_ov:AF190274
seq_documentation_block:
LOCUS AF190274 1327 bp mRNA linear VRT 06-OCT-1999
DEFINITION Bothrops jararaca carboxypeptidase homolog mRNA, complete cds.
ACCESSION AF190274
VERSION AF190274.1 GI:6013462
KEYWORDS
SOURCE jararaca.
ORGANISM Bothrops jararaca

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Lepidosteidae; Squamata; Scleroglossa; Serpentes; Colubroidae; Viperidae; Crocodylia; Bothrops.

REFERENCE
1 (bases 1 to 1327)
Murdoch, A.F., Hayashi, M.A.F. and Camargo, A.C.M.
Screening of Bothrops jararaca pancreas cDNA library
JOURNAL
Unpublished
2 (bases 1 to 1327)
Murdoch, A.F., Hayashi, M.A.F. and Camargo, A.C.M.
Direct Submision
Submitted (28-SEP-1999) Biochemistry, Butantan, Av. Vital Brasil, 1500, Sao Paulo, SP 05503-900, Brazil
LOCATION/Qualifiers
1.1327
/organism="Bothrops jararaca"
/db_xref="taxon:8724"
21.1271
/product="carboxypeptidase homolog"
/protein_id="AAF01344.1"
/translation="MPEFLIGATSAFAETTVHREDEKRYRTPRNEDEYVFLNYL
ANIVQDEWRPDSVEIVKAEIVDEFRIDEADRCSEYESILOSGLYEILIDLOAVLD
RQDNHARTAGYNEKYNSMEKIDAMTADIANENPISVSRLOIGTFRRPRLKVC
KGVNKKATFIDCGFPHAREMISPAFCOMPVRAVPTCKETITPOLKLDYITLPLV
NIDGVYSKOSRMGRKRSVAGSTCIGTDPRNPDAMCSVGSRNPCSTYCGSK
PSEKETRLADFTIRNRSIIOAYLTHSYQMLLPYSTYDLISNNKLSIKAEK
IRELVFETETYPGAAITTPAAGSGDDMAYDQIGVAFELDRKGRYFALPES
QIKPCEETMAVKKIAYIAEMLSY"

BASE COUNT 416 a 245 c 274 g 392 t
ORIGIN

alignment_scores:
Quality: 700.50 Length: 369
Ratio: 2.780 Gaps: 5
Percent Similarity: 68.293 Percent Identity: 38.482

alignment_block:
US-09-980-881-2 x AF190274 ..

Align seg 1/1 to: AF190274 from: 1 to: 1327

```

15 PheCysGluGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaI 31
   |||:||||| ||| |||:||||| |||:|||||
60 TTGCGAAGAACACAGCTTCACAGATTGATGGGAAAGGTTACCGGTG 109
   |||:||||| ||| |||:||||| |||:|||||
31 AleuProAlaGlyHisSerArgGlnValGlnValLeuGlnAsnLeuThrTh 48
   |||:||||| ||| |||:||||| |||:|||||
110 TACACCTCGCAATGAGATGATATATTTCTCAATTACTTACCCAAATA 159
   |||:||||| ||| |||:||||| |||:|||||
48 hTyrGluLeuValLeuThrGlnProValThrAlaSerLeuValLeuLys 64
   |||:||||| ||| |||:||||| |||:|||||
160 TTGTGCAAGCTGACTTTTGGCGCCAGATTCTGTGAACCTAGCAAGCA 209
   |||:||||| ||| |||:||||| |||:|||||
65 LysLysGlnValHisPhePheValAsnAlaSerAspValAlaSerValLys 81
   |||:||||| ||| |||:||||| |||:|||||
210 GAATGACGTGTGATTTCCGAATTTGAGACAGACAGTCTTCAAGTTGA 259
   |||:||||| ||| |||:||||| |||:|||||
81 SalenHisLeuAsnValSerGlyLysProCysSerValLeuLeuAlaAspV 98
   |||:||||| ||| |||:||||| |||:|||||
260 GAGCATTTCTACAGCAAGAGCTGATATATGAAATTTAATGATTAATC 309
   |||:||||| ||| |||:||||| |||:|||||
98 aGluAspLeuLeuGlnGlnGlnHisSerArgAsnAspThrValSerProA 114
   |||:||||| ||| |||:||||| |||:|||||
310 TTCAGAGCTCTCTTGTATGACCACTGGACAACTCACGCT.....CG 350
   |||:||||| ||| |||:||||| |||:|||||
115 AlaSerAlaSerTyrTyrGlnGlnTyrHisSerLeuAsnGluLeuTyrSe 131
   |||:||||| ||| |||:||||| |||:|||||
351 ACCGCTGATACATATGAAATAATATATAGTTGGAAATAATGATGAGC 400
   |||:||||| ||| |||:||||| |||:|||||
131 rTripleGluPheLeuThrGluArgHisProAspMetLeuThrLysIleH 148
   |||:||||| ||| |||:||||| |||:|||||

```

```

401 TTGACAGCTGATATGTCCAATGAAATCCAAATGCTTTGTTCCGATTAC 450
   |||:||||| ||| |||:||||| |||:|||||
148 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
   |||:||||| ||| |||:||||| |||:|||||
451 AAATAGGAACAACCTTGAAGGAGCGTCCAAAGCCTCTCCGAGAGGT... 497
   |||:||||| ||| |||:||||| |||:|||||
165 GlyLysGluGlnThrAlaLysAsnAlaIleTripleLysPcysGlyIleH 181
   |||:||||| ||| |||:||||| |||:|||||
498 GGAAGAGCTGTGTAAATTAATAAAGCTATCTTTATAGCTGTGGCTTCCA 547
   |||:||||| ||| |||:||||| |||:|||||
181 SalArgGluThrPheSerProAlaPheCysLeuThrPheIle..... 195
   |||:||||| ||| |||:||||| |||:|||||
548 TGCAGAGAAATGATCTACCTGCTTTTGCAGTGGTTGTACGAGAGG 597
   |||:||||| ||| |||:||||| |||:|||||
195 ..... 195
598 CTGTGCACCTTATGAAAGAAACATTTATGACTGACTTGTGAATTAAG 647
   |||:||||| ||| |||:||||| |||:|||||
195 ..... 195
648 TTAGACTTCTATATCTTGCTGTCTTGAATATGACGGTTATGCTTATTC 697
   |||:||||| ||| |||:||||| |||:|||||
196 ...GlyHisAsnArgMetLysPheValGlyAsnArgSerPheTyrAlaAsn 211
   |||:||||| ||| |||:||||| |||:|||||
698 TTGAAACAGAGTCGCATGTGAGAGAACGCTCTGTAAATGCTGTA 747
   |||:||||| ||| |||:||||| |||:|||||
211 snHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrp 227
   |||:||||| ||| |||:||||| |||:|||||
748 GTACATCATGTTGTTACTGATCCCAACAGAAATTT...GATGTCGATGG 794
   |||:||||| ||| |||:||||| |||:|||||
228 CysGluGluGlyAlaSerSerSerSerSerGlyLysTyrCysGlyLe 244
   |||:||||| ||| |||:||||| |||:|||||
795 TGCTGTGTTGGTCTTCAGAAATCCCTGATGAAACCTATTTGGTGAG 844
   |||:||||| ||| |||:||||| |||:|||||
244 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 261
   |||:||||| ||| |||:||||| |||:|||||
845 TAAACCTGAATCAGAAAGAGACCAAGCTTTGGCTGACTTCATTCGAA 894
   |||:||||| ||| |||:||||| |||:|||||
261 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 277
   |||:||||| ||| |||:||||| |||:|||||
895 GAAACCGTCATTTATCCAGGCTTACTTGAACCTATTTCTTACCTCCAG 944
   |||:||||| ||| |||:||||| |||:|||||
278 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 294
   |||:||||| ||| |||:||||| |||:|||||
945 ANGCTCCTTATCCATACCTTATCTTATGATCTAACATCAACAATTA 994
   |||:||||| ||| |||:||||| |||:|||||
294 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleAspLysHis 311
   |||:||||| ||| |||:||||| |||:|||||
995 AAACGTAATTTCTATGTGCTAAAGAACCATTCGAGAACTA...AAGTAC 1041
   |||:||||| ||| |||:||||| |||:|||||
311 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 327
   |||:||||| ||| |||:||||| |||:|||||
1042 TGTGGCGACCGAATATAGTATGTCGAGAGCTGCAACAAATTTATCT 1091
   |||:||||| ||| |||:||||| |||:|||||
328 AlaProGlyGlyLysAspArgPheLysPheLysPheLysLysTyrSe 344
   |||:||||| ||| |||:||||| |||:|||||
1092 GCTGTGTGGTCTCAGATGATGGCTTATGACAGGATTAATAATATAGC 1141
   |||:||||| ||| |||:||||| |||:|||||
344 rPheThr 346
   |||:||||| ||| |||:||||| |||:|||||
1142 CTTCAC 1148

```

seq_name: gb:om:SSC133775

seq_documentation_block:

LOCUS SSC133775 1251 bp mRNA linear MAR 09-JUL-1999

DEFINITION Sus scrofa mRNA for pancreatic procarboxypeptidase B.

ACCESSION AJ133775

VERSION AJ133775.1 GI:5457421

KEYWORDS carboxypeptidase B.

SOURCE pig.

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

seq_name: gb-dr:HS224866
seq_documentation_block:
LOCUS      HSA224866                1310 bp    mRNA    linear    PRI 12-MAR-1998
DEFINITION Homo sapiens mRNA for pancreatic procarboxypeptidase B.
ACCESSION  AJ224866
VERSION    AJ224866.1 GI:2959342
KEYWORDS   carboxypeptidase B.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 1310)
AUTHORS    Aviles, F.
TITLE       Direct Submission
JOURNAL     Submitted (05-MAR-1998) Aviles F., Bioquimica i Biologia Molecular,
MEDLINE     Universitat Autònoma de Barcelona, Cerdanyola 08193, Cerdanyola
FEATURES    08193, SPAIN
            2 (bases 1 to 1310)
            Aloy, P., Caldasus, L., Villegas, V., Reverter, D., Vendrell, J. and
            Aviles, F. X.
            Comparative analysis of the sequences and three-dimensional models
            of human procarboxypeptidases A1, A2 and B
            J Biol. Chem. 379 (2), 149-155 (1998)
            98182241
FEATURES    Location/Qualifiers
            1..1310
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /tissue_type="pancreas"
            5..50
            6..1259
            /EC_number="3.4.17.2"
            /function="zymogen of pancreatic carboxypeptidase B"
            /note="inactive precursor of carboxypeptidase B"
            /codon_start=1
            /evidence=experimental
            /product="procarboxypeptidase B"
            /protein_id="CAA12163.1"
            /db_xref="GI:2960072"
            /db_xref="SMIS-PROT:P15086"
            /translation="MALLLVLTVALASAHGGEHEGEKVRVNEDEHHIIRREL
            ASTQIDEMKPDSTYQIKPSTVDERKAEEDTVYENELKQNEQYKYLINLRNVE
            AOPDSRVATGHSYKRYKMETIETQVATEPNPALSIVIGTEGRAIVYLKVG
            KAGQKPAIFMDCGFHAREWISPAFCOMFVEAVRTGRETQVETLKKDFEYLPVL
            NIDGIYTWKSRFRKRTSTHTGSSCIGDPNPNRFDGMCIEASRNPCDETCGPA
            AESKEKALADPIRNKLSIKAVLTINYSOMMIIPSYAVKIGENNAEFLALAKAT
            VKELASLHGTYTGPGATITYPAAAGSDMAVDGGRIRYSPTFELRDIGRGFLPES
            QIRATCEFTFLAIKYVASVLEHLY"
            51..1256
            /product="procarboxypeptidase B"
BASE COUNT 353 a      316 c      305 g      336 t
ORIGIN
mat_peptide
/product="procarboxypeptidase B"
BASE COUNT 353 a      316 c      305 g      336 t
ORIGIN
Alignment_scores:
Quality: 641.00      Length: 382
Ratio: 2.514      Gaps: 7
Percent Similarity: 66.754      Percent Identity: 35.340
Alignment_block:
US-09-980-881-2 x HSA224866 ..
Align seg 1/1 to: HSA224866 from: 1 to: 1310
3 LeuGysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnI 19
::: ::::: ||||| ::::: |||
6 ANGTTGGCAGCTCTGGTCTGTGAGCTGGCCCTGGCATCTGCATCA 55
::: ::::: ||||| ::::: |||
19 sValPheAla...PheGlnSerGlyGlnValLeuAlaIleLeuProArgT 35
| ::::: ||||| ::::: |||
56 TGGTGGTGAGCACTTTGAAGCGAGAAAGGTTCGTTGTTAAGTTGAAG 105

```

```

35 hrSerArgGlnValGlnValLeuGlnIleuThrThrTyrGluIle 51
::: ::::: ||||| ::::: |||
106 ATGAAATACATTAACCTAATCCGCGAGTTGGCCAGACACCAATP 155
::: ::::: ||||| ::::: |||
52 ValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGlnVal 68
::: ::::: ||||| ::::: |||
156 GACTTGTGAAGCCAGATTCTGTCAACAATCAAACTCACTACACTAGT 205
::: ::::: ||||| ::::: |||
68 LHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisLeu 85
| ::: ||| ::::: ||||| ::::: |||
206 TGACTTCGTTTAAAGCAAGAACTACTGCACCTGGAGCAAGTTCTTAA 255
::: ::::: ||||| ::::: |||
85 snValSerGlyIleProCysSerValLeuValAspValGluAspLeu 101
::: ::::: ||||| ::::: |||
256 AGCAGATGAACTACATACAGTACTGATTAAGCAACCTGAGAAATGTG 305
::: ::::: ||||| ::::: |||
102 IlleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAla 118
::: ::::: ||||| ::::: |||
306 GTGGAGGCTCAGTTGTGATACCGGGTT.....CGTCACACAGACA 346
::: ::::: ||||| ::::: |||
118 rTyrTrpGluGlnIleThrHisSerLeuAsnGluIleTyrSerTrpIle 135
::: ::::: ||||| ::::: |||
347 CAGTTATGAGAAAGTCAACAAGTGGAAACGATAGAGGCTTGACCTCAC 396
::: ::::: ||||| ::::: |||
135 heIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySer 151
::: ::::: ||||| ::::: |||
397 AAGTCCGCACTGAGAAATCCAGCCCTCATCTCCGCACTGTTATCGGAAC 446
::: ::::: ||||| ::::: |||
152 SerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGlu 168
::: ::::: ||||| ::::: |||
447 ACATTGAGGAGGACCGCTATTACCTCTGAAAGTT...GGCAAACTGTG 493
::: ::::: ||||| ::::: |||
168 nThrAlaLysAsnAlaIleThrPheLysPheGlyIleHisIleAlaArgGlu 185
::: ::::: ||||| ::::: |||
494 ACAAAATAAACCTGCCATTTTCATGGACGTGGTTCCATGCCAGAGAGT 543
::: ::::: ||||| ::::: |||
185 rPheSerProAlaPheCysLeuThrPheIle..... 195
::: ::::: ||||| ::::: |||
544 GGAATTCTCTCGATCTCTGCCAGTGGTTGTAGAGAGGCTGTCTGATCC 593
::: ::::: ||||| ::::: |||
195 ..... 195
594 TATGACGTGAGATCCAACTGACAGACTTCTCCAACTAGACTTTTA 643
::: ::::: ||||| ::::: |||
196 ..... 196
644 TGCCCGCCTCTGCTCAATATGATGCTGCTACATCTACCTGGACCAAGA 693
::: ::::: ||||| ::::: |||
198 snArgMetIleThrArgLysAsnArgSerPheTyrAlaAsnHisCysIle 214
::: ::::: ||||| ::::: |||
694 GCCGATTTTGAAGAAAGACTCGCTCCACCATCTGATCTGATGATGAT 743
::: ::::: ||||| ::::: |||
215 GlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluGlu 231
::: ::::: ||||| ::::: |||
744 GGCACAGACCCCAACAGAAATTTT...GATGCTGTGGTGTGAATTTG 790
::: ::::: ||||| ::::: |||
231 yAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGln 248
::: ::::: ||||| ::::: |||
791 AGCCTCTGAAACCCCTGTGATGAACTACTGTGACCTGCCGACAGCT 840
::: ::::: ||||| ::::: |||
248 erGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
::: ::::: ||||| ::::: |||
841 CTGAAAGAGACCAAGCCCTGGCTGATTTGATCCGCAACAACTCTCT 890
::: ::::: ||||| ::::: |||
265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 281
::: ::::: ||||| ::::: |||
891 TCCATCAAGGCATATCTGACAAATCCACTGCTGCTCCAAATGATATGTA 940
::: ::::: ||||| ::::: |||
281 eProTyrSerTyrThrArgSerLysSerLysAspHisGluLeuSer 298
::: ::::: ||||| ::::: |||
941 CCCTTACTCATATGCTTACAAACTCGGTGAGAAACAATGCTGAGTTGATG 990
::: ::::: ||||| ::::: |||
298 euValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsnThr 314

```

```

991 CCTGGCTAAGCTACTGTAAGAACT...GCCCTACGACAGCGAC 1037
315 ArgTyrThrIstgIyHstgIySerGluThrLeuTyrLeuAlaProGly 331
1038 AAGTACACATATGCGCCCGGAGCTACACACATATCTCTGCTGGGG 1087
331 yGlyAspAspTyrPileTyrAspleuglyIleYstYrSerPheThr 346
1088 CTCGACGACTGGGCTATGACCAAGATCATATCTCTTCCAC 1133

```

seq_name: gb.pr:BC015338

seq_documentation_block:

```

LOCUS      BC015338          1454 bp    mRNA    linear    PRI 04-OCT-2001
DEFINITION Homo sapiens, clone MGC:21282 IMAGE:4422620, mRNA, complete cds.
ACCESSION  BC015338
VERSION    BC015338.1  GI:15929838
KEYWORDS   MGC.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1454)
            Strausberg, R.
            Direct Submission
            Submitted (01-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
            NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxill.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.

```

REMARK
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAX Plate: 27 Row: k Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4503002.

FEATURES

source

```

1..1454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:21282 IMAGE:4422620"
/tissue_type="Adrenal cortex, carcinoma"
/clone_11b="NIH_MGC_84"
/lab_host="DH10B"
/notes="Vector: PCMV-SPORT6"
5..1258
/codon_start=1
/product="Unknown (protein for MGC:21282)"
/protein_id="AAH1538.1"
/db_xref="GI:15929839"
/translacion="MALVLYTVALASAHGHEPEGEKVRVNVEDNHNITREL
ASTQIDFWKPSVTQIKPHSTVDPRVRAEDVTYENVIKQNELQYKVLISLRNVE
AOFDSRVATGHSYERKYNKWTIEAMTQVATENALTRSTIGTFEBERAYILKVG
KAGQKPAIFMPCGFHARBMISPARCOMFVRRAVRTGREGIVTELDLDFVLPVL
NIDGTYTWTKSRFWKTRSTGTSSCIGTDNRNPDACWCEIGASRNPCDETYCGPA
ASEKETALADPFINKLSSIKAYLLTISYSQMMIYPYSYAKLESNNAELNALAKAT
VEELASLHGKTYGPGATTIYPAGGSDMAVDOGIRYSFTFELDLGTGRYGFILPES
QIRATCEEFLEAIKRYVAVYLEHLV"

```

CDS

```

BASE COUNT      398 a      335 c      327 g      394 t
ORIGIN

```

alignment_scores:
Quality: 641.00 Length: 382
Ratio: 2.514 Gaps: 7
Percent Similarity: 66.754 Percent Identity: 35.340

alignment_block:
US-09-980-881-2 x BC015338 ..

Align seg 1/1 to: BC015338 from: 1 to: 1454

```

3 LeucylSerLeuAlaValLeuValProIleValLeuPheCysGluGlnI 19
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
5 ATGTGGACACTCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTA 54
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
19 sValPheAla...PheGlnSerGlyGlnValLeuAlaLeuProArgT 35
  |::::::::::::::::::::::::::::::::::::::::::::::::::::
55 TGGTGTGAGCAGCTTTGAAGCGAGAAAGTGTCTCGGTACCTTGAG 104
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
35 hrSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGluIle 51
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
105 ATGAATAATCATATACATATATCGCGAGTGTGGCAGCAGCAGCAGAT 154
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
52 ValLeuTyrGlnProValThrAlaAspleuIleValIstYstYsGln 68
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
155 GACTCTGGAAGCCAGATTCGTCTCACAACAATCAACCTCAGTACAGT 204
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
68 IHisPhePheValAsnAlaSerAspValAspAsnValIstYstYsLe 85
  ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
205 TGACTTCCTGTAAAGCAGAAATCTCTCAGTGTGAGAGTGTCTTAA 254
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
85 snValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 101
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
255 AGCAGAAATGAACTAATACAAAGTACTGATTAACCAACCTGAGAAAT 304
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
102 ILeuGlnGlnIleSerAsnAspThrValSerProAlaIleAlaSer 118
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
305 GTGAGAGCTCAGTTTGAAAGCCGGGT.....CGTCAGAACAGCACA 345
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
118 rTyrGluGlnIleThrIstYstYsLeuAsnGluIleTyrSerPheGlu 135
  ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
346 CAGTTATGAGAGTACACAAAGTGGAGAAAGTGAAGCTTGAGCTCAAC 395
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
135 heIleThrGluArgHisProAspMetLeuThrIstYsIleHisIleG 151
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
396 AAGTCGACAGCAGAAATCCAGCTCATCTCGCAGGTATATGGAAAC 445
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
152 SerPheGluIstYrProLeuTyrValLeuIstYsValSerGlyIstY 168
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
446 ACATTGAGGAGCGCGGTATTTACCTCTCGTGAAGGT.....GCCAA 492
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
168 nThrAlaIstYsAlaIleThrPheAspCysGlyIleHisIstYsVal 185
  ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
493 ACAAAATTAAGCTCCATTTTCATGAGACTGTGGTTCCATGCCAGAG 542
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
185 rPheSerProAlaPheCysLeuTyrPheIle..... 195
  ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
543 GGATTTCTCTGCAATTCGCAAGTGGTTTGAAGAGAGCTGTCTGAC 592
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
195 ..... 195
593 TATGAGCTGAGATTCAGAGTACAGAGCTTCTGCAAGATTAGACTTT 642
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
196 .....GLYHis.....A 198
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
643 TGTCTCTGCTGTGCTCAATTAATGAGCTCATATCTACACCTGACCA 692
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
198 snArgMetTyrArgIstYsAsnArgSerPheTyrAlaAsnAsnHisCys 214
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
693 GCCGATTTTGAAGAAAGCTGCTGCAACCATATGATGATGATGATG 742
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
215 GlyThrAspLeuAsnSerAsnPheValSerIstYsHisIstYsGluG 231
  ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```


128 CAGATTCGTGCACAAATCAAACTCAAGTCAAGTTCAGTTCCTCCGTG 177
73 AsnAlaSerPValAspAsnValLysAlaHisLeuAsnValSerGlyI 89
178 AAACCAAGAGATATATTTGGCTGTGGAGAGACTTCTTGACAGAAATGAAGT 227
89 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 106
228 ACATATGAGGTACTCATAAACAACTGAGATCTGTCTCTGAGAGCTAGT 277
106 LeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
278 TTGCACACAGAGTCT.....CGTCAACTGGACACAGATTATGAGAG 318
123 TyrHisSerLeuAsnGluIleTyrSerThrIleGlnPheLeuArgAla 139
319 TACACAACTGGGAAACGATCGAGGCTTGACTTGAAGAACTGACCAAGTGA 368
139 GHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluYst 156
369 AAATCCAGACCTCATCTCTGCAACAGCCATCGGAACATCACTTTTGGAA 418
156 YrProLeuTyrValLeuLysValSerGlyGluGlnThrAlaLysAsn 172
419 ACAATATATACCTCTCAAGATT.....GGCAACCTGGACCAAAATGAAGCT 465
173 AlaIleThrIleAspCysGlyIleHisAlaArgAlaGluThrIleSerProAl 189
466 GCATTTTCATGAGCACTGTGTTCATGCACGAGAAATGGATTTCCCATGC 515
189 aPheCysLeuThrPheLe..... 195
516 ATTTTGGCAAGTGGTGTGTGAGAGAGGCTGTCTCACTATGATATGAGA 565
195 195
566 GTCACATGACAGAAATTCCTCAACAACTAGACTTATATGTCTTCCTGTG 615
196GlyHis.....AsnArgMetTrpAr 202
616 CTCATATATTGATGGCTACATCTACACCTGGACCAAGCAAGATGGAG 665
202 GlyAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyTyrAspLeuAla 219
666 AAAGACCCCTGTACCAATGCTGGAACCTGCACTGATGGCACAGAGCCCA 715
219 snSerAsnPheValSerLysHisTyrCysGluGluGlyAlaSerSerSer 235
716 ACAGAAATTTT...GATGCTGGGTGGTCACAACTGGAGCCCTCTACAGAC 762
236 SerCysSerGlnThrTyrCysGlyLeuTyrProGlnSerGlnProGluVal 252
763 CCTCGGATGAGACTTACGTGATGATCTGCTGCAAGATCTGAAAAAGAGAC 812
252 LysAlaValAlaSerPheLeuArgTyrGlnAsnIleAsnGlnIleLysAlaT 269
813 CAAGGCGCTGGGTGATTTATACGCAACAACTCTCTCTCATCAAAAGCAT 862
269 yTrIleSerMetHisSerTyrSerGlnHisIleValAlaPheProTyrSerTyr 285
863 ACCTGACGATCCACTATCTACAGATATCTACTACCCCTATATCTAT 912
286 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGln 302
913 GATTCACAACTCCCGAGAACAAATGCTGAGTTGAATTAACCTGCTGAAGGC 962
302 uAlaValAlaGluAlaIleAspLysThrSerLysAsnThrArgTyrThrHisG 319
963 TCCCTGTAAAGAACTT...GCTACACTGATGACCAAGTACACATACG 1009
319 LysHisGlySerGluHisLeuTyrLeuAlaIleProGlyGlyGlyLysAspTrp 335
1010 GCCCAGAGACTTCAACAATATCTATCTCTGCTGCTGGGGGCTGTGATACAGG 1059

336 IleTyrAspLeuGlyIleLysTyrSerPheThr 346
|||||
1060 GCTTATGACCAAGGAAATCAATATTCCTCACC 1092

This Page Blank (uspto)

OM of: US-09-980-881-2 to: N_Geneseq_032802:* out_format : pfs

Date: Sep 18, 2002 7:02 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed-p2n.model -DEV=x1h
-O=/cgn2_1/USPRO_pool/US09980881/runat_16092002_140130_10714/app-query.fasta.1.1126
-DB=N_Geneseq_032802 -OPMT=fastep -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPT=0.000 -LOOEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=ptc
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pts
-NORTEXT=HEARSITE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 -CGN1_1.0 -NCPV=6 -ICPV=3 -LONGLOG
-DEV=TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-2
Query length: 360
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 386.310000

```

score_list:
Sequence      Strd Orig      ZScore      EScore Len  Documentation
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC81962 + 1911.00 3685.66 6.0e-197 15
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV74302 + 1832.00 3534.43 1.6e-188 12
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV11671 + 1832.00 3530.75 2.6e-188 17
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62846 + 1832.00 3530.75 2.6e-188 17
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ41001 + 1827.00 3521.03 8.9e-188 17
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV18005 + 889.00 1698.91 2.8e-86 14
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAV90600 + 640.50 1217.15 1.9e-59 12
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV42494 + 639.00 1213.78 2.9e-59 12
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV41795 + 639.00 1213.78 2.9e-59 12
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV42506 + 631.50 1199.00 1.9e-58 12
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV17331 + 623.00 1176.49 3.5e-57 21
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV1159.86 611.50 1159.86 2.9e-56 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV35760 + 604.00 1149.27 1.1e-55 92
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAQ09601 + 576.00 1094.88 1.2e-52 92
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV3839 + 564.00 1061.85 8.5e-51 7
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV76477 + 563.00 1067.28 4.2e-51 7
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV76476 + 563.00 1067.28 4.2e-51 7
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV76475 + 563.00 1065.24 5.5e-51 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV76478 + 563.00 1065.24 5.5e-51 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV81728 + 563.00 1061.68 8.7e-51 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV42500 + 561.00 1064.76 5.8e-51 99
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV42497 + 561.00 1064.53 6.3e-51 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV24804 + 554.00 1050.47 3.7e-50 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV42511 + 554.00 1050.47 3.7e-50 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62790 + 548.00 1038.80 1.6e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62801 + 548.00 1038.80 1.6e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62789 + 547.00 1036.85 2.1e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62791 + 546.00 1034.91 2.7e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV05477 + 546.00 1033.46 3.2e-49 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62805 + 545.00 1032.96 3.4e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62803 + 545.00 1032.96 3.4e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62804 + 545.00 1032.96 3.4e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62808 + 543.00 1029.07 5.7e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62800 + 543.00 1029.07 5.7e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV17322 + 543.00 1022.50 1.3e-48 18
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV41819 + 543.00 1022.50 1.3e-48 18

```

```

/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62802 + 542.00 1027.13 7.3e-49
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV62807 + 540.00 1023.23 1.2e-48
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV03838 + 528.50 1000.96 2.1e-47
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAV17587 + 519.50 981.35 2.6e-46
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV05476 + 519.50 980.89 2.7e-46

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC81962

seq_documentation_block:

```

ID AAC81962 standard; cDNA; 1573 BP.
XX
AAC81962:
XX
01-MAR-2001 (first entry)
XX
Human brain carboxypeptidase B cDNA.
XX
Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
XX treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
XX cerebroprotective; antialzheimers; neurotrophic;
XX hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
XX Down's syndrome; head trauma; ss.
XX
Homo sapiens.
XX
Key Location/Qualifiers
XX CDS 18..1100
XX FT /*tag= a
XX FT /product= "carboxypeptidase B"
XX
XX WC20006617-11.
XX
09-NOV-2000.
XX
01-MAY-2000; 2000MO-JP02878.
XX
30-APR-1999; 99UP-0125169.
XX
(MATS/) MATSUMOTO A.
XX
Matsumoto A;
XX WPI; 2000-687534/67.
XX P-PSDB; AAB11457.
XX
Human brain carboxypeptidase B isolated from the hippocampus useful for
XX screening agents for the treatment of Alzheimer's and other brain
XX disorders -
XX
Claim 2b; Page 64-68; 84pp; Japanese.
XX
This invention describes a novel protein with peptidase activity
XX against brain beta-amyloid precursor protein which has been isolated from
XX human hippocampus and which has cerebroprotective, antialzheimers,
XX neurotrophic, neuroprotective and hemostatic activity and which can be used
XX as a vaccine or for gene therapy. The protein, and compounds identified
XX by screening as promoters or inhibitors of its activity, are used to
XX regulate beta-amyloid accumulation in the brain and treat or prevent
XX diseases in which this occurs, such as Alzheimer's, senile dementia,
XX inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;
XX

```

alignment_scores:

Quality: 1911.00 Length: 360
Ratio: 5.308 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-980-881-2 x AAC81962 ..
Align seg 1/1 to: AAC81962 from: 1 to: 1573

```

1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCysG1 17
  |||
18 ATGAAGCTTTCAGCCCTGACGCTCTGTACCCATTGCTCTCTCTGTA 67
  |||
21 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
  |||
68 GCAGCATGCTCTCGGTTCCAGAGGGCCAGTCTAGCTGCTCTCTTA 117
  |||
34 rGthSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrGlu 50
  |||
118 GAACCTTAGGCAAGTTCAGTCTACAGAACTTACTACAACTATGAG 167
  |||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
  |||
168 ATTGTTCTCTGGCAGCCGCTACAGCTGACCTTATGGAAGAAAACA 217
  |||
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 84
  |||
218 AGTCCATTTTGTGTAATGATCATGTGATGCAATGTGAAAGCCATT 267
  |||
84 euAsnValSerGlyIleProCysSerValLeuAlaAspValGluAsp 100
  |||
268 TAAATGTGAGGAGGAATTCATGAGTCTGTGTCGACAGAGTGAAGAT 317
  |||
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
  |||
318 CTTATTACACGACGATTTCCACGACACAGTCACCCCGAGCCTCGC 367
  |||
117 aSerTyTyGluGlnTyHisSerLeuAsnGluIleTySerTrpIleG 134
  |||
368 ATGTAATAAGACAGATACACTCACTAAATGAATCTATCTTGATAG 417
  |||
134 LuPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleG 150
  |||
418 AATTATTAATGAGAGGATCTGATATGCTTACAAAAATCCACATGGA 467
  |||
151 SerSerPheGluLysTyProLeuTyValLeuLysValSerGlyLysG 167
  |||
468 TCCATCATTTAGAAAGTCCACCTGATGTTTAAAGSTTTCTGGAAGA 517
  |||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 184
  |||
518 ACAAAAGCCCAAAATGCCATATGATGATGATGATGATGATGATG 567
  |||
184 LuTrpIleSerProAlaPheCysLeuTrpPheIleGlnHisAsnArg 200
  |||
568 AATGGATCTCTCTGCTTCTGCTGTGTTGATAGGCCATTAATGCAAT 617
  |||
201 TrpArgLysAsnArgSerPheTyAlaAsnAsnHisCysIleGlyTh 217
  |||
618 TGGAGAAAGAACGTTCTTTCTATGCGAACATTCATTCGGAACAGA 667
  |||
217 PleuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAla 234
  |||
668 CCGAATATACAACTTGTCTCCAAACACTGCTGTGAGGAAGGTCATCA 717
  |||
234 SerSerSerCysSerGluThrTyCysGlyLeuTyProGlnSerGluPro 250
  |||
718 GTTCCCATGCTCGGAACCTACTGTGACTTATCTTAGGTAGAACCA 767
  |||
251 GluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIle 267
  |||
768 GAAGTAGAAGCAGTGGCTAGTTCTTGAGAGAAATATCACACAGATTAA 817
  |||
267 ValAlaTyIleSerMetHisSerTySerGlnHisIleValPheProTyr 284
  |||
818 ACATATACATCAGCATGATACATCCACGATATAGTGTTCATAT 867
  |||
284 SerTyThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
  |||
868 CCTATATACGAGATTAAGCAAAAGACATGAGAACTGTCTTAGAGCC 917
  |||

```

```

301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTy 317
  |||
918 AGTGAGAGAGTTGCTGCTATTTGACAAAACCTAGTAAATACAGATAT 967
  |||
317 rHisGlyHisGlySerGluThrLeuTyLeuAlaProGlyGlyAspA 334
  |||
968 ACATGGCCATGCTCAGAAAACCTTATACCTAGCTCTGAGAGTGGGAGC 1017
  |||
334 sPTripIleTyAspLeuGlyIleLysTySerPheThrSerAsnPro 350
  |||
1018 ATTGATCTATAGATTGGGATGCAATATTCGTTTACATCAAAACCCAC 1067
  |||
351 ValGluLysLeuLeuProLeuSerLeuLys 360
  |||
1068 GTAGAGAAAGCTTTTGGCGCTGTCTTAA 1097
  |||

```

seq_name: /STD1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV74302

seq_documentation_block:

ID AAV74302 standard; cDNA; 1272 BP.

AC AAV74302;

DT 28-APR-1999 (first entry)

DE Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.

KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;

KW polymorphism detection; thrombotic disease; ds.

OS Homo sapiens.

PN W09855645-A1.

PD 10-DEC-1998.

XX 02-JUN-1998; .98WO-EP03244.

XX 03-JUN-1997; 97US-0869057.

XX (SCHD) SCHERING AG.

PA Morser MJ, Nagashima M.

DR WPI: 1999-045800/04.

DR P-PSDB: AAW92270.

PT Detecting new polymorphism of human plasma carboxypeptidase B -
 PT comprises Alanine or Threonine at position 147 of protein by DNA or
 PT protein analysis, useful to detect risk of thrombotic disease in
 PT humans

XX Example 1; Page 24; 35pp; English.

XX This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
 XX hPCPBthr147. The invention relates to a method for determining the
 CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
 CC comprises obtaining a prepared tissue or blood sample and determining the
 CC presence of DNA coding for naturally occurring polymorphs of the protein
 CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
 CC respectively). Determination of the relative distribution of the PCPB
 CC polymorphs in a patient's blood by genetic or protein analysis by the
 CC methods is useful to determine the risk of thrombotic disease in humans.
 CC Such assessments may be made by accumulating information concerning the
 CC relative distribution of the different polymorphs within the general
 CC population compared with populations known to be at risk and establishing
 CC a PCPB polymorph profile for at-risk patients.

SO Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

alignment_scores:

Quality: 1832.00

Length: 415

Ratio: 5.103 Gaps: 3
Percent Similarity: 86.506 Percent Identity: 85.783

alignment_block:
US-09-980-881-2 x AAV74302 ..

Align seg 1/1 to: AAV74302 from: 1 to: 1272

```

1 MetIleuGysSerIleuAlaValIleuValProIleValIleuPheCysG1 17
1 ATGAAAGCTTGGACGCTTGGACGCTTGTGAACCATGCTCTCTCTGGA 50
17 uGlnHisValPheAlaPheGlnSerGlyGlnValIleuAlaIleuProA 34
51 GCAGCAGTCTTCGCTTTCAGATGGCCAAAGTTCAGCTGCTCTCTCTA 100
34 rGhrSerArgGlnValGlnValIleuGlnAsnIleuThrThrThrTyGlu 50
101 GAACTCTAGGCAAGTTCAGATGCTTACAGAACTTACTACAACTATAGAG 150
51 IleValIleuThrPglProValThrAlaAspLeuIleValIleuLysLysG1 67
151 ATGTCTCTGGCAGCGGTAACAGCTGACCTTATGTGAAGAAAAAACA 200
67 nValHisPhePheValAsnAlaSerAspValAspAsnValIleuHisL 84
201 AGTCATATTTTGTAAATGATCATCTGATGTCAGCAATGTAAGAACCAT 250
84 euAsnValSerGlyIleProCysSerValIleuAlaAspValGluAsp 100
251 TAAATGGAGCGGAAATTCATGAGCTGTCTGCGCAGCAGGTGGAAGAT 300
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
301 CTATTCAGACAGCAGATTTCCAGCAGACAGTCCGCCGAGCTCGGC 350
117 aSerTyTyGluGlnIleThrHisSerIleuAsnGluIleTySerTrpIleG 134
351 ATGCTACTATGAAACAGTATCATCTCACTAAATGAAATCTATCTGTGATAG 400
134 IupheIleThrGluArgHisProAspMetLeuThrIleHisIleGly 150
401 AATTTAAATCTGAGAGCATCTCGATATGCTTACAAAATCCACATTGGA 450
151 SerSerPheGluLysTyTyProLeuTyValIleuLysValSerGlyLysG1 167
451 TCCGCAATTTAGAAAGTACCCACTATGTTTAAAGTTCTCGAAAAAGA 500
167 uGlnThrAlaLysAsnAlaIleThrPheAspCysGlyIleHisAlaArgG 184
501 ACAAAACGCCAAAATGCAATATGATGATGACTGTGGAATCCATGCCAGAG 550
184 IuTrpIleSerProAlaPheCysLeuThrPheIleGlyHis..... 197
551 AATGGATCTCTCTGCTTCTGCTTGCTTGCTTCAATGAGCCATATACCTCAA 600
197 ..... 197
601 TTCTATGGATAATAGGCAATATACCAATCTCTGAGGCTGTGGATTT 650
197 ..... 197
651 CTATGTTATCCGGTGGTTAATGTGACGGTTATGACTACTCATGAGAAA 700
198 ..AsnArgMetTrpArgLysAsnArgSerPheTyAlaAsnAsnHisCys 213
701 AGAATCGAATGTGGAGAAAGACCGTTCTTTCTATGCGAACAATCATTTCC 750
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisIleTyCysGluG1 230
751 ATCGAACAAGACCTGATATAGGAAGCTTGTCCCAACACACTGTGTGAGA 800
230 uGlyAlaSerSerSerCysSerGluThrTyCysGlyLeuTyTrpG 247

```

```

|||||
801 AGTGCAATCCAGTTCTCATGCTCGGAAACCTACTGTGAGACTTATCCTG 850
247 IuSerGluProGluValIleValAlaValAlaSerPheLeuArgArgAsnIle 263
851 AGTCAGAACCCAGAGTGAAGCAGTGGCTGACTTCTTGAGAGAAATATC 900
264 AsnGlnIleLysAlaTyIleSerMetHisSerTySerGlnHisIleVa 280
901 AACCAAGTTAAAGCATCATGATGATGATCATATCTCCAGCATATAGT 950
280 IpheProTySerTyThrArgSerLysSerLysAspHisGluGluLeuS 297
951 GTTTCATATTCATACACAGAAATAAAGCAAGACCATGAGAACCTGT 1000
297 eLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsn 313
1001 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTTGAGAAACTAGTAATAT 1050
314 ThrArgTyThrHisGlyHisGlySerGluThrLeuTyIleuAlaProG1 330
1051 ACCAGGTATACATGAGCAGTGGCTCAGAAACCTTATACCTAGCTCTGG 1100
330 yGlyGlyAspAspTrpIleTyAspLeuGlyIleLysTySer...Phe... 345
1101 AGTGGGAGCAGATTGATCATGATTTGGCCATCAAAATATCGTTACAA 1150
346 .....Th 346
1151 TTGAACTTCAGAAATACGGGACCATACGAGATTCTTGCCGAGCCGTAC 1200
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
1201 ATCAAAACCACCTGTAGAGAGCTTTGCCGCTGTCTCTATAAA 1243
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA196.DAT:AA11671
seq_documentation_block:
ID AAT11671 standard; DNA: 1749 BP.
NC AAT11671:
XX
XX 12-APR-1996 (first entry)
DE Human plasma carboxypeptidase B coding sequence.
XX
XX KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
XX purification; plasminogen; affinity column; ss.
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX CDS 41..1312
XX FT /tag= a
XX FT /product= Human plasma carboxypeptidase B.
XX FT sig_peptide 41..106
XX FT /tag= b
XX FT mat_peptide 107..1309
XX FT /tag= c
XX
XX US5474901-A.
XX PN
XX PD 12-DEC-1995.
XX
XX PF 01-FEB-1991. 91US-0649591.
XX
XX PR 01-FEB-1991. 91US-0649591.
XX PR 14-OCT-1992. 92US-0959944.
XX PR 15-DEC-1993. 93US-0167727.
XX PR 19-JUL-1994. 94US-0277540.
XX
XX PA (GENTECH ) GENENTECH INC.
XX

```

PI Drayna DT, Eaton DL:
 XX MPI: 1996-039508/04.
 DR P-PSDB; AAR90293.
 XX
 PT Antibody to human plasma carboxypeptidase B - useful for detecting
 PT haemophilia A
 PS
 XX Disclosure; Figure 4; 40pp; English.
 CC
 CC (hPCPB) and does not cross react with other carboxypeptidases B
 CC useful for the detection of hPCPB in vitro. The antibody is also
 CC used for purifying hPCPB from a sample. Purification comprises
 CC passing a sample thought to contain hPCPB over either a column to
 CC which antibody has been bound, or a plasmaogen affinity column,
 CC eluting the column and then recovering the fraction containing the
 CC hPCPB.
 XX
 SQ Sequence 1749 BP: 521 A; 361 C; 342 G; 525 T; 0 other:

alignment_scores:
 Quality: 1832.00 Length: 415
 Ratio: 5.103 Gaps: 3
 Percent Similarity: 86.506 Percent Identity: 85.783

alignment_block:
 US-09-980-881-2 x AAT11671 ..

Align seg 1/1 to: AAT11671 from: 1 to: 1749

```

1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCysGI 17
41 ATGACCTTTGGCAGCTTGCAGTCTTGACCATGTTCTTCTGCTGA 90
17 uGlnHISValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
91 GCAGCATGCTTGGCGTTTCAGAGTGCCAAAGTTCAGCTCTTCCTA 140
34 rGThSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrGlu 50
141 GAACCTCTAGGCAAGTTCAGATCTTACAGATCTTACTACACATGTGAG 190
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGI 67
191 ATTGTTCTGTGGCAGCGGTGAACAGCTGACTTATGTGAAGAAAAACA 240
67 nValHISphePheValAsnAlaSerAspValAspAsnValLysAlaHISL 84
241 AGTCCATTTTGTAAATGCAATCTGATGCGCAATGTGAAGCCCATTT 290
84 euAsnValSerGlyIleProCySerValLeuLeuAlaAspValGIuAsP 100
291 TTAATGCTGAGCGAATTCATGCAAGTCTGTGCGACAGCGTGAAGAT 340
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
341 CTTATTCAACAGCAGATTTCCAAACGACACAGTCCGCGCCGCGCTCCC 390
117 aSerTyTyTyGlnGlnIleThrHISerLeuAsnGluIleTySerTrpIleG 134
391 ATCGTACTATGAAAGTATCACTACTAAATGAATCTATTTTGGATAG 440
134 LuPheIleThrGluArgHISProAspMetLeuThrLysIleHISLleGly 150
441 AATTATATACTGAGAGCATCTGATATGCTTAACAATAATCAACATTTGA 490
151 SerSerPheGlnLysTyTyProLeuTyValLeuLysValSerGlyLysGI 167
491 TCCTCATTTGGAGAGTACCACCTCTATGTTTTTAAAGGTTTCTGAAAAAGA 540

```

```

167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHISAlaArgS 184
541 ACAAAACAGCCAAAATGCAATGATGATTGACTGTGAATCATGACAGAG 590
184 LuTrpIleSerProAlaPheCysLeuTrpPheIleGlnHIS 197
591 AATGATCTCTCCGCTTCTGCTGTGCTTACAGCCATATACACAA 640
197 ..... 197
641 TTCTATGGATTAATAGGCAATATACCAATCTCTGAGCTTGTGATTT 690
197 ..... 197
691 CTATGTTATCCCGTGGTTAATGTCAGCGTTTATGCTACTATGAAAA 740
198 .. AsnArgMetTrpArgLysAsnArgSerPheTyAlaAsnAsnHISCyS 213
741 AGAATCGAATGTGAGAAAGAACCGTTCTTATGCGAAACATCATTCG 790
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHISTrpCysGluGI 230
791 ATCGAACAAGACCTGAATAGAACTTGTCTCCAAACACTGTGTGAGGA 840
230 uGlyAlaSerSerSerSerCysSerGluThrTyTyCysGlyLeuTyTyProG 247
841 AGGTGATCATCAATCTCTCATGTCGGAACCTATGTCGACTTTATCCG 890
247 LuSerLupProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
891 AGTCAAACCCAGAAAGTGAAGGAGTGTGCTATGTTCTTGAAGAAATATC 940
264 AsnGlnIleLysAlaIleTrpIleSerMetHISerTySerGlnHISLLeVa 280
941 AACCAATTAAGACATACATCACTGACATGCAATTCATCTCCACCATATAT 990
280 LpheProTySerTyTyThrArgSerLysSerLysAspHISGluLys 297
991 GTTTCATATTCCTTATACAGAAATGAAGTAAAGCAAGCATGAGAACTGT 1040
297 erLeuValAlaSerGluAlaValAlaArgAlaIleAspLysThrSerLysAsn 313
1041 CTCTGATACCGAGTGAAGCAGTTCGTCTATGTGAAGAAACCTAAGTAAAT 1090
314 ThrArgTyTyThrHISGlyHISGlySerGluThrLeuTyTyLeuAlaProGI 330
1091 ACCAGGTATACATAGGCCATGCGTCAGAAACCTTATACCTTACCTCTGG 1140
330 uGlyGlyAspAspTrpIleTyTyAspLeuGlyLleLysTyTySer...Phe... 345
1141 AGGTGGGAGCATTTGATCTATGATTTGGCATCAAAATATTCGTTACAA 1190
346 .....Th 346
1191 TTGAACCTCGAGATACGGGACATACGGAATCTTCTGCGGAGCGTTAC 1240
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
1241 ATCAAAACCACTGTAGAGAAAGCTTTGCGCTGTCTTAAAA 1283

```

seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA1997.DAT:AA162846
 seq_documentation_block:
 ID AA162846 standard; DNA: 1749 BP.
 XX AA162846;
 AC
 XX
 XX
 DT 08-NA1-1997 (first entry)
 XX
 XX Human plasma carboxypeptidase B coding sequence.
 DE Human plasma carboxypeptidase B; PCPB; haemostatic regulation;
 KW plasma; plasmaogen; ss.

```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 41..1312
XX FT /tag= a
XX FT /product= Human PCPB
XX FT sig_peptide 41..106
XX FT /tag= b
XX FT mat_peptide 107..1309
XX FT /tag= c
XX FT misc_binding 134..177
XX FT /tag= d
XX FT /bound_molecly= 46_bp_probe
XX PN US5593674-A.
XX PD 14-JAN-1997.
XX PF 01-FEB-1991; 91US-0649591.
XX PR 01-FEB-1991; 91US-0649591.
XX PR 14-OCT-1992; 92US-0959944.
XX PR 15-DEC-1993; 93US-0167727.
XX PR 19-JUL-1994; 94US-0277540.
XX PR 27-APR-1995; 95US-0430787.
XX PA (GENE) GENENTECH INC.
XX PI Drayna DT, Eaton DL;
XX DR WPI; 1997-099413/09.
XX DR P-PSDB; AAM14733.
XX XX
XX PT Using human plasma carboxypeptidase B in blood coagulation - is
XX PT functionally related to carboxypeptidase A and pancreas
XX PT carboxypeptidase B
XX PS Example 2; Column 37-42; 39pp; English.
XX CC This sequence encodes human plasma carboxypeptidase B (PCPB) which
XX CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
XX CC PCPB may be used therapeutically in hemostatic regulation. PCPB is
XX CC purified from human plasma or by transformed cell culture by
XX CC extraction using plasminogen bound to a solid phase.
XX SO Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

alignment_scores:
Quality: 1832.00 Length: 415
Ratio: 5.103 Gaps: 3
Percent Similarity: 86.506 Percent Identity: 85.783

alignment_block:
US-09-980-881-2 x AAT62846 ..
Align seg 1/1 to: AAT62846 from: 1 to: 1749

1 MetTysLeuCySserLeuAlaValLeuValProTleValLeuPheCysG1 17
|||||
41 ATGAGAGCTTGGACCTTGACAGTCTGTGACCAATGTTCTTCTGTGGA 90
|||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
|||||
91 GCACAGATCTCTCCGCTTTCAGAGTGGCCAAATCTACGCTCTTCTTA 140
|||||
34 rGThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGlu 50
|||||
141 GAACCTCTAGGCAAGTCAAGTCTACAGAACTTACTACAAATATAG 190
|||||
51 lIleValLeuTrpGlnProValThrAlaAspLeuValLysLysLysG1 67
|||||

```

```

191 ATTGTTCTCTGGACCGGTAAACAGCTGACCTTATGTCGAAGAAAAACA 240
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
|||||
241 AGTCATTTTTTTGTAAATGCATCTGATGTGCACAAATGTGAAGCCCAT 290
|||||
84 euAsnValSerGlyTleProCySerValLeuLeuAlaAspValGluAsp 100
|||||
291 TAAATGTAGCGGAATTCATCATGCTGCTGCTGGACAGCTGGAGAT 340
|||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaGlnSer 117
|||||
341 CTATTTCAACAGCAGATTTCACAGACAGTCCAGCCCGAGCTCCGCG 390
|||||
117 aSerTyrTrpGlnGlnTyrHisSerLeuAsnGluTleTyrSerTrpIleG 134
|||||
391 ATCGTACTATGACAGATTCACCTCACTAAATCAATCTATCTTGGAAT 440
|||||
134 lupheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
|||||
441 AATTATATACTGAGAGCATCTCGATATGCTTACAAAATCCACATTGGA 490
|||||
151 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysG 167
|||||
491 TCCCTATTGAGAAAGTACCACCTCTATGTTTAAAGTTTCTGGAAGA 540
|||||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
|||||
541 ACAAAACAGCAAAAATGCCATATGATGATGACGTGGAAATCCATGCCAG 590
|||||
184 lUTrPLeSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
|||||
591 AATGATCTCTCCCTCTTCTGCTGCTGATGATGAGCAATTAATCA 640
|||||
197 ..... 197
641 TTCATATGGAAATATAGGCAATATACCAATCTCTGAGCGTTGTGAT 690
|||||
197 ..... 197
691 CTATGTTATGCCGGGTGTTAATGTGAGCGTTATGACATCATATGAAAA 740
|||||
198 ..AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCys 213
|||||
741 AGAATCGAATGTGAGAAAGAACCGTCTTCTGATGCGAACAATCATGTC 790
|||||
214 lIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGlu 230
|||||
791 ATCGGAACAGACCTGAATAGCAACTTTCCTTCCAAACACTGGCTGAG 840
|||||
230 uGlnAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyrProG 247
|||||
841 AGGTCAATCCAGTTCCTCATGCTCGGAAACCTACGTGTGACATTATCT 890
|||||
247 lUsSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 263
|||||
891 AGTCAGAACCAAGAGTGAAGCGAGCTAGCTTCTTGAGAGAAATATC 940
|||||
264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIle 280
|||||
941 AACCAATTAAGCATACATACAGCATTCATCTCCACCAATATAGT 990
|||||
280 lPheProTyrSerTyrThrArgSerLysSerLysAspHisGluLeuS 297
|||||
991 GTTTCATATTCCTTATACAGAAATGAAAGCAAAACCATGAGAACTGT 1040
|||||
297 eTLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsn 313
|||||
1041 CTCTAGTAGCCAGTAGACAGAGTCTGCTATGAGAAAATCTGTAATAAT 1090
|||||
314 ThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProG 330
|||||
1091 ACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGG 1140
|||||

```

```

330 yGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer... 345
|||||
1141 AGGTGGGACGACATGGATCTATGATTTGGGCATCAATATTCGTTTACAA 1190
|||||
346 ..... 346
1191 TTGAACCTTCGAGATACGGGACATACGAGATTCTGCTGCCGAGCGTTAC 1240
|||||
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
|||||
1241 ATCAAAACCCACCTGTAGAGAGCTTTTCCGCTGCTCTAATAA 1283
|||||
seq_name: /STDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ41001
seq_documentation_block:
ID AAQ41001 standard; cDNA; 1749 BP.
XX
AC AAQ41001;
XX
XX 24-AUG-1993 (first entry)
DE Human plasma carboxypeptidase B gene.
XX
XX PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX sig_peptide 41..106
XX mat_peptide 107..1312
XX /tag= a
XX /tag= b
XX /product= PCPB
XX /tag= c
XX /function= probe
XX /note= "used to obtain full-length clones"
XX
XX US5206161-A.
XX
XX 27-APR-1993.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI: 1993-151724/18.
XX P-PSDB; AAR36273.
XX
XX New human plasma carboxypeptidase B - used as haemostatic
XX regulator for clotting blood, partic. for treating blood clotting
XX disorders, e.g. haemophilia
XX
XX Disclosure: fig 4; 40pp; English.
XX
XX Human plasma carboxypeptidase B was isolated from human plasma and
XX partially sequenced. Oligonucleotide primers were designed based on
XX the partial amino acid sequences. The primers were used in a PCR
XX amplification to identify cDNA encoding PCPB from a human liver cDNA
XX library. The PCR product was capable of encoding the first 37 amino
XX acids of PCPB; a 46mer probe was used to obtain the full-length
XX sequence which, although disclosed in the specification, is not
XX claimed. PCPB inhibits the enzymatic conversion by tPA of
XX plasminogen to plasmin in the presence of fibrinogen.
XX
XX Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

```

```

alignment_scores:
  Quality: 1827.00      Length: 415
  Ratio: 5.103
  Percent Similarity: 86.265      Percent Identity: 85.542
alignment_block:
US-09-980-881-2 x AAQ41001
Align seg 1/1 to: AAQ41001 from: 1 to: 1749
1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCysG1 17
|||||
41 ATGAAGCTTTCACACCTTGCAGTCTTGTACCCATTCTCTCTCTCTGTA 90
|||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuPro 34
|||||
91 GCACATGCTCTTCCGCTTTTCAAGAGTGGCAGAGTTCAGTCTCTCTCT 140
|||||
34 rGThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGlu 50
|||||
141 GAACCTCTAGGCAAGTTCAGTCTTACAGAACTTCTACTACAACTATGAG 190
|||||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
|||||
191 ATGTCTCTGGCAGCCGCTTACAGCTGACCTTATTTGTGAAGAAAAACA 240
|||||
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHis 84
|||||
241 AGTCCATTTTTCATTAATCATCTGATCTGCGCAATGTGAAGCCCAT 290
|||||
84 euAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
|||||
291 TAAATGAGCGGAAATTCATGCAGTGTCTGCTGCAGACGTAAGAT 340
|||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
341 CTATATTCACAGCAGATTTCCAAACAGACAGTCCGCCGCCCTCCGC 390
|||||
117 aSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleG 134
|||||
391 ATGTCATTAAGAACAGTATCCTACTAATATGAATCTATCTTGGATG 440
|||||
134 IuPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
|||||
441 AATTATTAACGTGAGAGCATCTCGATATCTTACAAAAATCCACATGGA 490
|||||
151 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysG1 167
|||||
491 TCCTCATTTGAGAAAGTACCCACTATATGTTTAAAGGTTCTGGAAAAA 540
|||||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgG 184
|||||
541 ACAACACGCCAAAAATGCCATATGATGATGACTGTGCAATCCATGCCAGG 590
|||||
184 IuTrpIleSerProAlaPheCysLeuThrPheIleGlyHis..... 197
|||||
591 AATGATCTCTCTGCTTCTGTTGTTATGAGGCATATTAATCAAA 640
|||||
197 ..... 197
641 TTCATATGGATATATAGGCAATATACCAATCTCCGAGGCTTGTGAAAT 690
|||||
197 ..... 197
691 CTAATGTTATCGCGGTGTTAATGTGACGGTTATGACTACATCATGAAAA 740
|||||
198 ..AsnArgMetThrArgLysAsnArgSerPheTyrAlaAsnHisCys 213
|||||
741 AGAATCGAATGTGAGAAAGAACCGTCTTCTTATGCAACATCATTCG 790
|||||
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluG1 230
|||||
791 ATCGAAGACGACCTGATATGAACTTCTTCCAAACACTGCTGTGAGGA 840

```



```

291 SASPHISGLUleuSerleuValAlaSerGluAlaValArgAlaIle 308
    |||||||
460 AGACCATGAGGAGACTGCTCTAGTACGACGAGACGATTCGCTATG 509
    |||||||
308 SPLSThrSerAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
    ::|||
510 AGAATACTAGTAAATACAGGTATACATGCGCATGGCTGAGAAAC 559
    |||||||
325 LeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyI 341
    |||||||
560 TTATACCTAGCTCCTGGAGGTGGGAGCATTTGATCTATGATTTGG 609
    |||||||
341 elvSTySer-Phe..... 345
    |||||||
610 CAATATTCCTTTCAATTGAACTTCAGATACGCGACATACGATTC 659
    |||||||
346 .....ThSerAsnProProValGluTyrSerleuProLe 357
    |||||||
660 TGCTGCGGAGCGCTTACATCAACACCTGTAGAGAACTTTGGCCGT 709
    |||||||
357 uSerleuLys 360
    |||||||
710 GTCTCTTAAA 719

```

seq_name: /sids1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:AAQ90600

seq_documentation_block:

ID AAQ90600 standard; DNA; 1215 BP.

AAQ90600;

13-MAR-1996 (first entry)

Porcine Tyr-His-Met procarboxypeptidase B coding sequence.

Procarboxypeptidase B:carboxypeptidase B; Pichia; pCPB;

human serum albumin; premating factor alpha; mating factor alpha;

procBB; da.

Sus scrofa.

Key Location/Qualifiers

CDS 1..1215

/*tag= a

/product= Porcine procarboxypeptidase B.

MO9514096-A1.

26-MAY-1995.

16-NOV-1994; 94WO-US13142.

16-NOV-1993; 93US-0153258.

(ELIL) LILLY & CO ELI.

Fayerman JT, Greenen DP, Hersherberger CL, Larson JL;

Sterner JL, Zhang H;

WPI: 1995-200386/26.

P-PSDB: AAR75131.

DNA encoding porcine carboxypeptidase B - used for transforming

host cells, partic. Pichia species, for prodn. of the enzyme

Claim 2; Page 20-21; 34pp; English.

The porcine carboxypeptidase B coding sequence can be placed in a

bacterial or pref. Pichia yeast expression vector. The expression

vector further comprises the signal peptide of either human serum

albumin (designated pUCD23 - NRRL B-21029); premating factor alpha

(designated pFJ489 - NRRL B-21028); mating factor alpha (designated

pFJ474 - NRRL B-21032) or the porcine procBB signal peptide,

CC (designated pUCD27 - NRRL B-21027). The method can be used for
 CC producing large amounts of porcine carboxypeptidase B and when
 CC produced in Pichia yeast, the protein does not need solubilisation
 CC or folding. The produced enzyme is then used for pref. cleaving
 CC basic residues from the carboxy terminus of proteins.
 CC
 XX
 SQ Sequence 1215 BP; 358 A; 300 C; 266 G; 291 T; 0 other;

alignment_scores:

Quality: 640.50 Length: 361
 Ratio: 2.680 Gaps: 6
 Percent Similarity: 66.205 Percent Identity: 37.673

alignment_block:

US-09-980-881-2 x AAQ90600 ..

Align seg 1/1 to: AAQ90600 from: 1 to: 1215

23 PheGlnSerGlyGlnValleuAlaIleuProArgThrSerArgGlnVa 39

28 TTCGAAGGGGAGAGAGGTTCCTGTCATGTGGAAGATGAATGACAT 77

39 IglValleuGlnAsnLeuThrThrTyrGluIleValleuTrpGlnP 56

78 CAGCTACTCTCATGAGTGTGGCCAGACAGGAGATGACTTCTGGAAC 127

56 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 72

128 CAGATTCCTGCACCAAAATCAACCTCAGACGTACGTGCTCGTGG 177

73 AsnAlaSerAspValAlaAspValLysAlaHisLeuAsnValSerGlyI 89

178 AAGCAGAGAGATTTTGGCTGTGGAAGACTTCTGACAGACATGACT 227

89 eProCysSerValleuLeuAlaAspValLysLysGlnValHisPhePheVal 106

228 ACAATATGAGTACTCATTAACACCTGAGATCTGCTGAGGCTCACT 277

106 IeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122

278 TGCACGACGAGCTC.....CGTACACTGACACAGTTATGAGAG 318

123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 139

319 TACAACAACTGGGAAACGATCGAGCTTGACTAGCAAGTACACAGTCA 368

139 gHisProAspMetLeuThrLysIleHisIleGlySerPheGluLysT 156

369 AATCCAGACCTCATCTCGCACAGCCATCGAATCATTTTATGAGAA 418

156 yrProLeuTyrValleuLysValSerGlyLysGluGlnThrAlaLysAsn 172

419 ACATATATATCTCTCTCAAGTTT...GGCAACCTGGAACCAATATGAGCT 465

173 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 189

466 GCCATTTTCATGACTGTGTTTCATGCGACAGAAATGATTTCCAGC 515

189 aPheCysLeuTrpPheIle..... 195

516 ATTTTGCCAGTGTGTTGTGAGAGAGGCTGTTCTACATGATGATAGAGA 565

195 195

566 GTCACATGACAGAAATTCCTCAACAAAGCTAGACTTTATGCTTCGCTTG 615

196GlyHis.....AsnArgMetTrpPar 202

616 CTCAATATATGATGGCTCATCTACACCTGACCAAGAACGAAATGTGAG 665

202 GlyAsnArgSerPheTyrAlaAsnHisCysIleIleLysThrAspLeu 219

```

666 AAGACCCGCTACCAATGCTGGAATACCTGATGCGACAGACCCCA 715
219 snSerAspHeValSerLysHisTrpCysGluGluAlaSerSerSer 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
716 ACAGAAATTTT...GATGCTGGGTGTCACAACTGAGCCCTTACAGAC 762
236 SerCysSerGluThrTrpCysGlyLeuTrpProGluSerGluProGluVa 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
763 CCGTCGATAGACTTACTGTGATCTGTCACAGAGTCGTGAAAAAGAGAC 812
252 lLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAla 269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
813 CAAGCCCTGCGTATTTTATACGCAACAACTCTCTCCATCAAGCAT 862
269 yLileSerMetHisSerTrpSerGlnHisIleValPheProTrpSerTrp 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
863 ACCTGACGATCACTCATCTACACAGATGATCTACCTTATCTCAT 912
286 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
913 GATTACAACTCCCGAGAACAACTGATGATGATACCTGGCTAAGGC 962
302 uAlaValArgAlaIleAspLysTrpSerLysAsnThrArgTrpThrHisG 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
963 TGGCGTAAAGAACTT...GCTACACTGTATGGCACCAGTACATACG 1009
319 lYHisGlySerGluThrLeuTrpLeuAlaProGlyGlyGlyAspAspTrp 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1010 GCCCAGAGGTACAAACAATCATCTGCTGCTGGGGCTGTGATGACGTGG 1059
336 lIleTrpAspLeuGlyIleLysTrpSerPheThr 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1060 GCTTATGACCAAGCAATCAATATTCCTTCACC 1092

seq_name: /SIDSI/gcgdata/geneseq/emb1/NA196.DAT:AA12494
seq_documentation_block:
ID AA12494 standard; DNA; 1263 BP.
XX
AC AA12494;
XX
DT 12-FEB-1997 (first entry)
XX
DE Human pancreatic carboxypeptidase B encoding sequence.
XX
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPPT;
KW mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Fd; F(ab')2; Fc; leader; human carboxypeptidase B; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1248
    /tag=a
    /product= preproenzyme
FT sig_peptide 1..39
    /tag=b
    /note="partial enzyme secretion leader sequence
    (pre-sequence); residues -108 to -96"
FT sig_peptide 40..324
    /tag=c
    /note="Pro-sequence; residues -95 to -1"
FT mat_peptide 385..1245
    /tag=d
    /note="mature enzyme; encodes residues +1 to +307"
XX
XX WO9620011-A1.
XX
XX 04-JUL-1996.
XX
XX 21-DEC-1995; 95WO-GB02991.

```

```

XX
PR 16-AUG-1995; 95GB-0016810.
PR 23-DEC-1994; 94GB-0026192.
XX
XX (ZENE ) ZENECA LTD.
XX
PI Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
PI Tarragona-Fiol A, Taylorson CJ;
XX
DR WPI: 1996-321650/32.
XX
PT P-PSDB; AAW06172.
XX
PT Two component system for anti-tumour therapy - comprising targeting
PT moiety linked to mutated enzyme which can transform an
XX anti-neoplastic prodrug
XX
PS Reference Example 14; Page 131; 182pp; English.
XX
XX A two-component system for anti-tumour therapy comprises a targeting
XX moiety linked to a mutated enzyme which can transform an anti-neoplastic
XX prodrug. The system is based on antibody directed enzyme prodrug therapy
XX (ADEPT) using non-naturally occurring mutant forms of host enzymes,
XX pref. human pancreatic ribonuclease (HP-RNase), (see AA12478-83).
XX Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
XX be used. The present sequence encodes the native prepro-HCPB (see
XX AAW06172) contained in vector pCII1698.
XX
SQ Sequence 1263 BP; 343 A; 306 C; 302 G; 312 T; 0 other;

alignment_scores:
    Quality: 639.00 Length: 379
    Ratio: 2.526 Gaps: 7
Percent Similarity: 66.755 Percent Identity: 35.620

alignment_block:
US-09-980-881-2 x AA12494 ..

Align seg 1/1 to: AA12494 from: 1 to: 1263

6 LeuAlaValIleValProIleValLeuPheCysGluGlnHisValPheAl 22
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4 CTCCTGGTTCGTGACTGCTGCGCCCTGGCATCTGCTCATGATGGTGGA 53
22 a...PheGlnSerGlyGlnValIleuAlaLeuProArgTrpSerArg 38
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
54 GCACCTTGAAGCGGAGAGAGTGTCCGTTAACCTGACATGAATAATC 103
38 lValAlaValIleuGlnAsnLeuThrTrpTrpTrpGluIleValLeuTrp 54
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 ACATTAAACATAATCCGCGAGTGGCCACAGACACCAATGATGACTCTGG 153
55 GlnProValThrAlaAspLeuIleValLysLysLysGlnValHisPhe 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 AAGCCAGATCTGTCCACACAATATCAACCTCAGTACAGTCACTTCG 203
71 eValAsnLaseAspValAspValIleValHisValHisValSerG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 TGTTAAGCAGAGACTGTCACCTGTGAGAAAGTTTAAAGCAGATG 253
88 lYleProCysSerValLeuLeuAlaAspValGluAspLeuIleGln 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 AACTATACATACAGTACTGATTAAGCAACGAGAAATGTGTGAGGCT 303
105 GlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTrpTrpG 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 CAGTTTGATACCGCGGT.....CGTCAACAGCAGACACAGTTTAA 344
121 uGlnTrpHisSerLeuAsnGluIleTrpSerTrpIleGluPheIleTrp 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 GAAGTACACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCCGA 394

```

```

138 LuArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 154
139 CTEGAGAAATCCAGCCCTCATCTCTGCACTGCTTAACTGGAAACCACTTTGAG 444
145 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLys 171
146 GAGCGGCTATTTCCTCTGTAAGTT...GGCAAGCTGGACAAATTA 491
171 GAGAAATLeuTyrLysPheGlyLysIleHisIleAlaArgLysThrPheSer 188
172 GCGCGCAATTCATGAGCTGCTTCCATGACAGAGAGATTTCTCTC 541
188 roAlaPheGlyLeuThrPheLe... 195
542 CTGCAATCTGCAAGTGGTTTGTAGAGAGCGCTTCTACTATGACGT 591
195 ..... 195
592 GAGATCCAGTGAAGAGCTTCTGCAAGTGTAGACTTTATGCTCTGCC 641
196 .....GlyHis.....AsnArgMet 201
642 TGTGCTCAATATTGATGCTATCATCTACCTGACCAAGACGCGATTTT 691
201 rPArGLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 217
692 GAGAAAGATCTGCTCCACCATCTGATCTAGCTTATGTCACAGAC 741
218 LeuAsnSerAsnPheValSerLysHisTyrCysGluGluLysAlaSer 234
742 CCCAAGAGAAATTTT...GATGCTGGTGGTGTGAATTTGAGACCTCTCG 788
234 rSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluPro 251
789 AAACCCCTGATGAAACTTACTGCTGACCTGCCGACAGACTCGAAAGC 838
251 LuValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLys 267
839 AGACCAAGGCCCTGCTGATTTATCCGCAACAACTCTCTTCCATCAG 888
268 AlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 284
889 GCATATCTGCAATCCATCTGCTGCTCCAAATGATGATCATCCCTTAC 938
284 rTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 301
939 ATATGCTTCAAACTCGGTGAGAAATGCTGAGTTGAATGCCCTGCTGA 988
301 rGluAlaValAlaArgAlaIleAspLysThrSerLysAsnThrArgTyr 317
989 AAGCTACTGTGAAGAACTT...GCCTCACTGACGCGACCAAGTACACA 1035
318 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAsp 334
1036 TATGGCCGCGGAGCTCAACAAATCTATCTGCTGGGCGGCTGTGACGA 1085
334 pTPrIleTyrAspLeuGlyLysLysTyrSerPheThr 346
1086 CTGGGCTTATGACCAAGATGATATCTCTGAC 1122
seq_name: /Sids1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:AAV41795
seq_documentation_block:
ID AAV41795 standard; DNA; 1263 BP.
XX
AC AAV41795:
XX
DT 20-NOV-1998 (first entry)
XX
DE Human pancreatic carboxypeptidase nucleotide sequence.
XX
KW ser: human; pancreatic carboxypeptidase B; insulin; protein sequencing;
prodng therapy.

```

```

XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT CDS 1..1248
FT /tag= a
FT /product= "pancreatic carboxypeptidase B"
FT /note= "No start codon given"
XX
PN W09835988-A1.
XX
PD 20-AUG-1998.
XX
PF 10-FEB-1998; 98WO-GB00415.
XX
PR 29-OCT-1997; 97GB-0022727.
PR 14-FEB-1997; 97GB-0003104.
PR 18-OCT-1997; 97GB-0022003.
XX
PA (ZENE ) ZENEPA (ZENE ) ZENECA LTD.
XX
PI Edge MD:
XX
DR WPT: 1998-467168/40.
DR P-PSDB: AAM74476.
XX
PT New modified pro-domain of carboxy-peptidase B - enhances expression
PT of co-expressed proteins for production of recombinant
PT carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT enzyme prodng therapy
XX
PS Example 1; Page 52-53; 83pp; English.
XX
CC The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC modified pro-domain of CPB on a separate gene to enhance recombinant
CC expression. This process can be used to produce recombinant CPB in
CC eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
CC antibody production and protein sequencing, while its fusions with
CC antibodies are useful in antibody-directed enzyme prodng therapy. The
CC modified pro-domain provide increased yields of recombinant CPB, possibly
CC by protecting the C-terminus against enzymatic degradation or by
CC increasing intracellular trafficking.
XX
SQ Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;
XX
alignment_scores:
Quality: 639.00 Length: 379
Ratio: 2.526 Gaps: 7
Percent Similarity: 66.755 Percent Identity: 35.620
XX
alignment_block:
US-09-980-881-2 x AAV41795 ..
Align seg 1/1 to: AAV41795 from: 1 to: 1263
6 LuValAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAl 22
||| ||||| ||||| ||||| |||||
4 CTCTGGTGTCTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53
22 a...PheGlnSerGlyGlnValLeuAlaValLeuProArgThrSerArg 38
||| ||||| ||||| ||||| |||||
54 GCACCTTTGAAGCGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 103
38 LuValGlnValLeuGlnAsnLeuThrThrThrTyrGluLysLeuValLeu 54
||||| ||||| ||||| ||||| |||||
104 ACATTAACTAATCCGCGAGTTGGCCAGACAGACCCAGATTTGACTCTGG 153
55 GlnProValIThrAlaAspLeuLysLysLysGlnValHisPhePhe 71
||||| ||||| ||||| ||||| |||||
154 AAGCAGATTTCTGTCACAAATCAAACTCAGATTAAGTTGACTCTGCC 203
71 eValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSer 88

```

```

204 TGTAAAGCAGAGATCTGCTACTGTGGAGATGTTCTTAAGCAGATG 253
88 LYLIEPROCYSSerValLeuLeuAlaAspValGluAspLeuIleGln 104
254 AACTACAAATCAAGTAGTGTATAGCAACCTGAAATGTGGTAGAGCT 303
105 GlnIleSerAsnAspThrValSerProAlaSerAlaSerTyrGln 121
304 CAGTTTCATAGCCGGGTT.....CTGCACACGACGACACTTATGA 344
121 uGlnTyrHisSerLeuAsnGlnIleTyrSerProIleGlnPheIleThr 138
345 GAAGTACAAACAGATGGAAACGATAGAGCTTGACTCAACACAGTCCCA 394
138 LIArHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 154
395 CTGGAATTCAGCCCTCATCTCTCGCAGTGTATCGAACCACACTTTGAG 444
155 LysTyrProLeuTyrValLeuLysValSerGlyLysGlnIleThrAla 171
445 GACGCGCCATTATTACCTCTGAAGGTT...GGCAACCTGGACAAATTA 491
171 sAanAlaIleTyrIleAspGlyIleHisAlaArgGluTyrIleSerP 188
492 GCCTGCATTTTCATGAGCTGTGTTCCATGACAGAGATGATTTCTC 541
188 roLAPheCysLeuTyrPheIle..... 195
542 CTGATTTCTGCCAGTGTGTTGTAGAGAGCTGTCTACTAATGAGAGT 591
195 ..... 195
592 GAGATCCAGTAGACAGACTTTCGACAGTAGACTTTATGTCCTGCC 641
196 ..... GlyHis ..... AsnArgMet 201
642 TGTGCTCAATATGATGCTACATCTACACCTGGACCAAGGCCGATTTT 691
201 rPArGlyAsnArGserPheTyrAlaAsnAsnHisCysIleGlyThrAsp 217
692 GGAGAAAGACTGCTCCACCATCTGATCTACTGATTCATGGACAGAC 741
218 LeuAsnSerAsnPheValSerLysHisTyrCysGlnIleGlyAlaSer 234
742 CCCAACGAAATTTT...GATGCTGTTGGTGTGAATTTGAAATGGCTCTCG 788
234 rSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluPro 251
789 AAACCCCTGTGATGAACCTTACTGTGAGCTGCCGACAGAGTCTGAAGA 838
251 LuValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIle 267
839 AAACCAAGGCCCTGCGATTTTCATCGCAACAACCTCTCCATCAAG 888
268 AlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 284
889 GCATATCTGCAATCCACTGCTACTCCCAATGATCATCTACCTTATCTC 938
284 rTyrThrAspSerLysSerLysAspHisGlnIleLeuSerLeuValAs 301
939 ATATGCTTACAACTCGTGAGAACATGCTGAGTTGAAATGCCCTGGCTA 988
301 erGlnAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThr 317
989 AAGTACTGTGAAGAAGACTT...GCCACACTGCAGGCCACACAAGTACA 1035
318 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAs 334
1036 TATGGCCCGGAGCTACACAACTATCTCTCTCGGGGGCTCTGACGA 1085
334 pTPRIleTyrAspLeuGlyIleLysTyrSerPheThr 346

```

```

1086 CTGGGCTTATGACCAAGACATCATATTCCTTACC 1122
seq_name: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42506
seq_documentation_block:
ID AAAT42506 standard; DNM: 1284 BP.
XX
AC AAAT42506;
XX
DT 17-FEB-1997 (first entry)
XX
DE PROHCPB gene with PelB leader sequence.
XX
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPEPT;
KW mustard-ribonuclease; antibody directed enzyme producing therapy;
KW anti-neoplastic; produg; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1275
FT sig_peptide 1..66
FT sig_peptide /note= "PelB leader sequence"
FT sig_peptide 67..351
FT /tag= c
FT /note= "pro sequence"
FT mat_peptide 352..1272
FT /tag= d
FT /note= "mature HCPB sequence"
XX
PN MO9620011-A1.
XX
PD 04-JUL-1996.
XX
PE 21-DEC-1995; 95MO-GB02991.
XX
PR 16-AUG-1995; 95GB-0016810.
PR 23-DEC-1994; 94GB-0026192.
XX
PA (ZENEC) ZENECA LTD.
XX
PI Blakey DC, Boyle FT, Davies DH, Eggelte HD, Heaton DW;
PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
PI Tarragona-Fiol A, Taylorson CJ;
XX
DR WPI; 1996-321650/32.
DR P-PSDB; AAW06175.
XX
PT Two component system for anti-tumour therapy - comprising targeting
PT moiety linked to mutated enzyme which can transform an
PT anti-neoplastic produg
XX
PS Reference Example 18; Page 140-142; 182pp; English.
XX
CC A two-component system for anti-tumour therapy comprises a targeting
CC moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC produg. The system is based on antibody directed enzyme producing therapy
CC (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC be used. The present sequence is a cloned pro-HCPB gene contained in
CC plasmid pIC11738 and which can be expressed in E. coli.
XX
SQ Sequence 1284 BP; 355 A; 315 C; 299 G; 315 T; 0 other;

```

alignment_scores:
Quality: 631.50
Ratio: 2.588

Length: 361
Gaps: 6

XX Claim 1: Page 30; 36pp; English.
 XX
 CC The present sequence is a cDNA encoding novel human protein (NHP),
 CC carboxypeptidase-related protein. The carboxypeptidase-related
 CC proteins share structural similarity with animal proteases, and
 CC especially carboxypeptidase B or carboxypeptidase A. The
 CC carboxypeptidase-related protein cDNA is useful for the detection of
 CC mutant human carboxypeptidases, or inappropriately expressed human
 CC carboxypeptidases for the diagnosis of disease, and also as a
 CC therapeutic. It is also useful in drug screening, for generation of
 CC antibodies, as reagents in diagnostic assays, for the identification of
 CC other cellular gene products related to human carboxypeptidases, and as
 CC reagents in assays for screening for compounds that are useful for
 CC treating mental, biological or medical disorders and diseases. Nucleotide
 CC constructs encoding functional NHPs can also be used in gene therapy.
 XX
 SQ Sequence 1311 BP; 391 A; 278 C; 284 G; 358 T; 0 other;

alignment_scores:
 Quality: 611.50 Length: 417
 Ratio: 2.427 Gaps: 12
 Percent Similarity: 60.432 Percent Identity: 33.813

alignment_block:

US-09-980-881-2 x AAD03837 ..

Align seg 1/1 to: AAD03837 from: 1 to: 1311

```

1 MetLysLeuGys.....SerLeuAlaValLeuValProIleValLeu 15
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
43 CTGCGCTTTTGGCTGGCTTTTGAAGATTCTGCACCG..... 81
15 eCysGluGlnHisValPhe.....AlaPheGlnSerGlyGlnValLeu 30
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
82 .GGGCACAGCACCTTATATACACCGCTATGCTGGTGATAAAGTGATA 130
30 LaIaLeuProArgThrSerArgGlnValGlnValLeuGlnAsnLeuThr 46
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
131 GATTATTTCCCAACAGACAGAGACATATGACATGCAAGAAATATATCC 180
47 ThrThrTyrGluIleValLeuThrGlnProValThrAlaAspLeuIle 63
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
181 TATCAACTTAAGGTGACCTGTGACCCAGACAGATATCTCCATATGATC 230
63 Lys.....LysLysGlnValHisPhePheValAsnAlaSerAspVal 78
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
231 AGAGGAGACAGTACTGATGATGATCCCAAAATGTTCCGAGCC. 279
78 spAsnValLysAlaHisLeuAsnValSerGlyIleProCysSerValLeu 94
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
280 ....CTGTAGGCTTCTTACAGAGAACCAACATCCAGTACAGGTCCTC 324
95 LeuAlaAspValAlaGluAspLeuIleGlnGlnIleSerAsnAspThrVal 111
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
325 ATAGAGATCTTCAGAAACACTGAGAGAGGAGAGAGCTTGACACCCA 374
111 IsSerProAlaSerAlaSer.....TyrTyrGluGlnIleTyrHisSer 126
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
375 GAGAAACCGAAGATCCCTCTGATATATATGAAATTATCACTCTCT 424
126 euAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAla 142
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
425 TAGAAGAAATTCAAAATTCATGATGATCATCATTAATAAAGCTCAGCC 474
143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeu 159
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
475 CTCATTCACATGTTCTCTATGAGATCATATAGGAGGAAGTCTTTT 524
159 rValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrp 176
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
525 TATTTTAAAGCTG...GGCAGACGATACGACTCAAAAGAGCTGTTGGA 571

```

```

176 leaSPcysGlyIleHisAlaArgGluTrpIleSerProIlePheCysLeu 192
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
572 TAGACTGTGGTATTCATGCAAGAAATGATGGTCTGCTGCTTTTGCAG 621
193 TrpPheIle.....GlyHisAsnArgMetTrpArgLysAsnArg 195
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
622 TGGTTTATAAAGAGCTCTTACATATATAGAGTACCCAGCATGAG 671
195 ..... 195
672 AAAAATTGAAATCATATATATTTATATCATGCTGTGTTAAGCTCG 721
196 .....GlyHisAsnArgMetTrpArgLysAsnArg 205
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
722 ATGGATACCATTTTACTGTGACCAATGATGATTTTGAGAAAAACAG 771
206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsn 222
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
772 TCAGAGAACTCAAGGTTCCGCTGCGGAGAGTGCATGCCAATGAAACTG 821
222 eValSerLysHisTrpCysGluGlnGluAlaSerSerSerCysSerG 239
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
822 GAAAGTGAAAG...TGGTGTATGAAAGAGCTTCTATGACACCTTGATG 868
239 LuThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaVal 255
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
869 ACAATATCTGTGCGCTTTCCAGAAATCTGAGCCGGAAGAGAGCTGTGA 918
256 AlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSer 272
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
919 GCTAACTTCCTTGAAACACAGAAACACATATGAGGCTTATCTCTCT 968
272 ThrSerTyrSerGlnHisIleValPheProTyrSerTyrThrArgSer 289
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
969 TCATGATATGCTCAGATGTTACTGTATCCCTATTTCTTCAAAATATGCA 1018
289 ysSerLysAsnHisGluLeuSerLeuValAlaSerGluAlaValArg 305
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1019 CAATTCCCAATTTTACATGTTGAGAAATCTGCACCTTATTAAGCTGTGANT 1068
306 AlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGly 322
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1069 GCACCTT...CAGTCAATATACAGGAGTACATACAGATATGAGCCCTC 1115
322 rGluThrLeuTyrLeuAlaProGlyGlyLysAspAspTrpIleTyrAsp 339
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1116 CACAACGTTGTATGATGAGCTGTGAGCTCATGATGATTTGGGCTTACAAA 1165
339 euGlyIleLysTyrSerPheThr.Ser..... 347
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1166 ATGGAATACCTTATGATTTGCTTTCGAACCTAGCAGACACTGGATATTTT 1215
348 .....AsnProProValGluLysLeu 354
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1216 GGAATTTTACTCCAGAGATGCTCATCAAAACCACTGTACGAAGATA 1264

```

seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA1996.DAT.AAT35760

seq_documentation_block:

ID AAT35760 standard; cDNA: 927 BP.

XX AAT35760;

XX 09-NOV-1996 (first entry)

XX Rat mature carboxypeptidase B cDNA.

XX Carboxypeptidase-B; pro-enzyme; protease; insulin; ss.

XX Rattus sp.

XX Location/Qualifiers

FH Key

carboxypeptidase A. The carboxypeptidase-related protein cDNA is useful for the detection of mutant human carboxypeptidases, or inappropriately expressed human carboxypeptidases for the diagnosis of disease, and also as a therapeutic. It is also useful in drug screening, for generation of antibodies, as reagents in diagnostic assays, for the identification of other cellular gene products related to human carboxypeptidases, and as reagents in assays for screening for compounds that are useful for treating mental, biological or medical disorders and diseases. Nucleotide constructs encoding functional NHs can also be used in gene therapy.

SQ Sequence 2128 BP; 640 A; 461 C; 436 G; 591 T; 0 other;

	alignment_scores:	
Quality:	564.00	Length: 462
Ratio:	2.220	Gaps: 15
Percent Similarity:	54.978	Percent Identity: 30.736

```
alignment_block:
```

Align seg 1/1 to: AAD03839 from: 1 to: 2128

```

400 ATATATATACAGTGGCCCTGGTATACGCTCCAAAGATATGCCACAGCTAAG 449
      ::::::::::::::::::::::::::::|||
21 eala.phegInserGlyValLeuAlaAlaLeuProArgThrSerrArg 37
      ||| ::::::::::::::::::::|
450 AACAGCATATCAGTGAAATAAGTGAATGATTATATCCCAAAACAGAAAG 459
      ::::::::::::::::::::::::::::|
38 GlnValGlnValLeuGlnAsnLeuThr_ThrThyArgGluLeval.... 52
      ::::::::::::::::::::|
500 GAAGCATATCAGCTGAAGAAATAATATCCATACACTTAAAGGTGGTTCCTG 549
      ::::::::::::::::::::::::::::|
53 .....LeuThrGlnProValThralAspLeuLevalys 64
      ||| ::::::::::::::::::::|
550 CACTACAGTGGAGCTGGTGGCAGCCACAGATGCTCCATGATGATACAG 559
      ::::::::::::::::::::::::::::|
65 .....LysLysGlnValHisPhePheValAsnAlaSerAspValAspHis 79
      ::::::::::::::::::::|
600 GGAACAGATTACTGCTGGTCCATATATCCCCCAAAATGGTCCCGAAGC.... 644
      ::::::::::::::::::::::::::::|
79 nValLysAlaHisLeuAsnValSerGlyLleProCysSerValLeuLeu 96
      ::| ||| ::::::::::::::::::::|
645 .CTGTTAGCCTTCTTACAGGAAGCCAAATCCAGTACAAAGTCTCTATAG 653
      ::::::::::::::::::::::::::::|
96 laAspValGluAspLeuIleGlnGlnGlnIleSerAsnAspThrValSer 112
      ::::::::::::::::::::|
694 AAGATCTTTCAGAAACAACCTGAGAAAGGAGACAGCTTGCACACCCAGAGA 743
      ::::::::::::::::::::|
113 ProArgAlaSerAlaSer....ThyArgGlnGlnThyHisSerLeuAs 127
      ||| ||| ::::::::::::::::::::|
744 AACCCAGAAATCCCTCTGTGATTAATTAATGAAGTTTAACTCCCTTGA 753
      ::::::::::::::::::::::::::::|
127 ngIuIleLysSerTrpIleGluPheIleThrGluArgHisProAspMet 144
      ||| ||| ::::::::::::::::::::|
794 AGAATTCCAAAATGAGATGCATCATCTGAATAATAACTCAGAGCCTCA 843
      ::::::::::::::::::::::::::::|
144 eutHrLysIleHisIleGlySerSerPheGluLysLysTrpLeuTrpVal 160
      ::| ::| ||| ||| ::::::::::::::::::::|
844 TTGCACATGTTCTCTATTTGGAAGATCATGATGAGAGGAAGATCTCTTTTAT 853
      ::::::::::::::::::::::::::::|
161 LeuLysValSerGlyLysGluGlnThrLysAsnAlaIleTrpIleAs 177
      ||| ||| ::::::::::::::::::::|
894 TTAAAGCTG...GGCAGACGATCCAGCACTCAAAAGAGCGTGGTAGTACA 940
      ::::::::::::::::::::::::::::|
177 pCysGlyIleHisAlaArgIuTrpIleSerProAlaPheCysLeuTrp 194
      ||| ||| ::::::::::::::::::::|
941 CTGGGATTCATGCACAGAGATGGAATGGCTGCTGCTTTGTGCAGATGCT 950
      ::::::::::::::::::::::::::::|
194 heile.....GlyHis_Asn..... 198
      |||:::| |||

```

991	TTGTAAAGAGAGTCTAGAAAAACAGAGCTCACAAATGTCAGAAATGTACT	1046
1398	198
1041	AAATTTACAAAATATCTCTGCCACTACCAAAAACAAAGATATGCTTAA	1090
1398	198
1091	TCCTTGAAGTATAGTAATAAAATTTTCTAAACATTCCTAAAAAGCTTT	1140
1398	198
1141	CTAACAATATAGAGTGACCCAGCCATGTGAAAAATGTTGAATCATCTATA	1190
1398	198
1191	TTTCTATATCATGCTGTGTAAAGTATGATGATACCATTTTGTGGA	1240
1399	199
1241	CCMATGATCGATTTTGGAGAAAAACMAGGTACAGMATCTCAAGTTTCGC	1290
213	CysIleGlyThrAspLeuAsnSerSerAsnIleValSerIleHisIleProCysGlu	229
1291	TECCGTGGAGTGGATGCCAAATAGAACTGGAAAGTAGAG...TGGTGTGA	1337
229	uGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyrP	246
1338	TGAGGAGAGCTTCTATGCACCTGTGTGATGACACATCTGTGGCCCTTTTC	1387
246	roGluSerGluProGluValIleValIleAlaSerPheLeuArgArgAsn	262
1388	CAGAATCTGAGCGGAGATGGAAGGCTGTAGCTACTTCCTCGAAAAACAC	1437
263	IleAsnGluIleValAlaTyrIleSerSerHisSerIleTyrSerGluHisIle	279
1438	ACAAAGCACATTAGGAGCTTATCTCTCTTCACTGATATAGCTCACAGATGT	1487
279	eValPheProTyrSerTyrThrArgSerIleSerIleAspHisGluIul	296
1488	ACTGATCCCTATCTTTCAAAATATGCAACAATCCCAATTTTGAATGTG	1537
296	euSerLeuValAlaSerGluAlaValArgAlaIleAspIleThrSerIleS	312
1538	TGGATCTGCAAGCTTATTAAGCTGTGATGCACCT...CACTCAGTATAC	1584
313	AsnThrArgTyrThrHisGlyHisIleSerGluIleThrIleuTyrIleAlaPr	329
1585	GGGGTAGCAATACAGATATAGGACCCAGCCCTCCACACAGTTGTATGTACGCTC	1634
329	oGlyIleGlyAlaAspArgTyrPheTyrAspLeuGlyIleIleTyrSerPheT	346
1635	TGGTATGCTCAATAGATTTGGGCTTACAAAAATAGGAATACCTTATGCATTGG	1684
346	hr.Ser.....	347
1685	CTTTGCAACTAGCTGACACTGGATATTTTGGATTTTCTACCCAGAGATG	1734
348AsnProProValGluIleLeu	354
1735	CTCATCAAAACCCACTGTACAGAAAACTA	1762

This Page Blank (uspto)

OM of: US-09-980-881-2 to: Issued_Patents_NA:* out_format : pfs
 Date: Sep 18, 2002 6:55 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
 -O=cg2_1/USPTO_spool/US09980881/rnmat_16092002_140129_10653/epc_query.fasta_1.1126
 -DB=Issued_Patents_NA -OEW=fastap -SUFFIX=emi -CAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
 -GAPOP=4.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -GAPOP=6.000 -DELEXT=7.000 -TGAPOP=10.000 -TGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62
 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ps
 -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=us09980881.ecg1.1.62 -NCP=6 -ICPU=3 -LONGLOC
 -DEV=TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPMX -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-2
 Query length: 360
 Database: Issued_Patents_NA:*
 Database sequences: 383533
 Database length: 122816752
 Search time (sec): 90.520000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation	..	5e-214	1272	1
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-869-057-1		1832.00	3976.37	8.5e-214	1272	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-07-649-591B-2		1832.00	3972.25	1.4e-213	1749	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-277-540-2		1832.00	3972.25	1.4e-213	1749	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-430-787A-2		1832.00	3972.25	1.4e-213	1749	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-696-139-1		640.50	1363.65	2.2e-68	1215	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-56		639.00	1361.85	3.6e-68	1263	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-71		631.50	1345.19	3.1e-67	1284	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-09-171-945-124		623.00	1319.87	7.9e-66	2154	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-782-760-5		604.00	1289.14	4.1e-64	927	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-696-139-3		576.00	1227.86	1.1e-60	921	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-67		561.00	1193.93	8.2e-59	999	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-54		561.00	1193.25	8.2e-59	1053	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-74		554.00	1177.84	6.4e-58	1059	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-77		553.00	1175.64	8.5e-58	1059	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-09-171-945-112		543.00	1146.37	3.6e-56	1870	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-640-906-3		513.00	1085.82	8.6e-53	1251	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-09-395-936-3		513.00	1085.82	8.6e-53	1251	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-640-906-1		487.50	1029.87	1.1e-49	1257	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-09-395-936-1		487.50	1029.87	1.1e-49	1257	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-09-331-709-2		213.00	438.04	1.0e-16	591	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-998-416-125		168.50	342.29	2.2e-11	515	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-484-105-15		118.50	212.25	0.0004	2504	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-484-106-15		118.50	212.25	0.0004	2504	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-782-760-2		100.50	200.92	0.0017	285	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-782-760-2		100.50	200.92	0.0017	285	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-706-216-1		95.00	159.68	0.3317	2719	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-609-049A-29		95.00	151.08	0.9993	5285	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-09-170-996-29		91.00	143.88	2.51	4680	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-254-358-1		91.00	143.88	2.51	4680	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-475-391-1		91.00	143.88	2.51	4680	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-709-609-1		90.50	150.04	0.3600	1333	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-09-073-297-1		86.00	148.54	1.38	1965	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-09-178-252-26		87.50	150.30	1.10	1576	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-689-974-2		87.50	150.30	1.10	1576	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-09-058-376-2		86.50	136.73	6.30	3797	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-07-915-203-1		86.50	136.73	6.30	3797	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-272-887-1		86.50	136.73	6.30	3797	1			
/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-789-449-1		86.50	136.73	6.30	3797	1			

/cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-609-049A-27 + 86.00 128.03 19.20 6831
 /cg2_6/pdata1/1/ina/5A.COMB.seq:US-09-170-996-27 + 86.00 128.03 19.20 6831
 /cg2_6/pdata1/1/ina/5A.COMB.seq:US-07-807-043B-9 - 85.00 132.27 11.15 4157
 /cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-299-849B-9 - 85.00 132.27 11.15 4157
 /cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-142-368A-9 - 85.00 132.27 11.15 4157

seq_name: /cg2_6/pdata1/1/ina/5A.COMB.seq:US-08-869-057-1

seq_documentation_block:

Sequence 1, Application US/08869057
 Patent No. 5985562
 GENERAL INFORMATION:
 APPLICANT: Morser, Michael J
 APPLICANT: Nagashima, Mariko
 TITLE OF INVENTION: Method of Detecting Thrombotic Disease
 TITLE OF INVENTION: Risk
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Berlex Biosciences Legal Department
 STREET: 15049 San Pablo Avenue
 CITY: Richmond
 STATE: California
 COUNTRY: USA
 ZIP: 94804-0099
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/869,057
 FILING DATE: 03-JUN-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Washlien, Wendy L.
 REGISTRATION NUMBER: 36,301
 REFERENCE/DOCKET NUMBER: 51509AUSM1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 510-262-5411
 TELEFAX: 510-262-7095
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1272 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 PUBLICATION INFORMATION:
 AUTHORS: Eaton, Dan L.
 AUTHORS: Malloy, Beth E.
 AUTHORS: Tsai, Siao P.
 AUTHORS: Hensel, William
 AUTHORS: Drayna, Dennis
 TITLE: Isolation, Molecular Cloning, and Partial
 Characterization of a No. 5985562el Carboxypeptidase B
 JOURNAL: J. Biol. Chem.
 VOLUME: 266
 ISSUE: 32
 PAGES: 21833-21838
 DATE: No. 5985562 15-1991
 US-08-869-057-1

alignment_scores:

Quality: 1832.00 Length: 415
 Ratio: 5.103 Gaps: 3
 Percent Similarity: 86.506 Percent Identity: 85.783

alignment_block:

US-09-980-881-2 x US-08-869-057-1 ..
 Align seg 1/1 to: US-08-869-057-1 from: 1 to: 1272

```

1 MetLysLeuCySerLeuAlaValProIleValLeuPheCysG1 17
1 ATGAAGCTTTTCAGCCCTTGACCTCTTACCCATGTTCTCTCTCTGA 50
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
51 GCAGCATCTCTTCGGCTTTCAGAGGGCCAAAGTTCTAGCTCTCTCTTA 100
34 TgThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrGly 50
101 GAACCTTAGGCAAGTTCAGTTCAGAGATCTTACTACAACTATGAG 150
51 IleValLeuThrProValThrAlaAspLeuIleValLysLysG1 67
151 ATTGTCTCTGGCAGCCGGTAACAGCTGACCTTATTGGAGAAAAACA 200
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHsL 84
201 AGTCATATTTTGTAAATGCAATCTGATGTCAGCAATGTGAAGCCCAT 250
84 euAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
251 TAAATGTGAGGAGAAATTCATGAGTGTCTGCTGGCAGAGTGGAAGAT 300
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
301 CTATATTCACAGCAGATTTCCACAGCACAGTCAGCCCGAGCCCTCGC 350
117 aSerTyTyTyGluGlnTyHisSerLeuAsnGluIleTySerThrIleG 134
351 ATGTAATCTAGAACAGTATCAGTACACATAAATGAATCTATCTTGATG 400
134 LurHeIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
401 AATTATTAATGAGAGCATCTGATATGCTTACAAAAATCCACATGGA 450
151 SerSerPheGluLysTyTyProLeuTyValLeuLysValSerGlyLysG1 167
451 TCCTCATTTGAGAAATACCCACTCTATGTTTAAAGGTTCTTGAAAGA 500
167 uGlnThrAlaLysAsnAlaIleThrPheAspCysGlyIleHisAlaArg 184
501 ACAAAAGCCCAAAATGCCATATGATGATGCTGGAATCCATGCCAGAG 550
184 LurPheLeuSerProAlaPheCysLeuThrPheIleGlyHis..... 197
551 AATGATCTCTCTGCTTCTGCTGTGTTGATAGGCCCATATATACCTCA 600
197 ..... 197
601 TTCTATGGGATATAGGCAATATACCAATCTTCCTGAGCCTTGATTT 650
197 ..... 197
651 CTATGTTATGCCGGTGTAAATGTGAGCGTTATGACTACTCATGAGAAA 700
198 ..AsnArgMetIrrArgLysAsnArgSerPheTyTyLalaAsnHisCys 213
701 ACAAATGCAATGTGAGAAAGAAACCGTCTTCTATGCGCAACATCATTC 750
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisIrrPcysGlu 230
751 ATCGGAACAGACCTGAATGGAACCTTGTCTCCAAACATGCTGTGAGAG 800
230 uGlyAlaSerSerSerSerSerSerGluThrTyTyCysGlyLeuTyProG 247
801 AGGTGCATCTCAATCTCATGCTCGAAACCTTACGTGACTTTATCTCG 850
247 LuSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
851 AGTCAGAACCAAGTGAAGGAGTGGCTAGTTTCTTGAGAAAGAAATATC 900

```

```

264 AsnGlnIleLysAlaTyIleSerMetHisSerTySerGlnHisIleVal 280
901 AACCAAGATTAAAGCATATACATGACATGATTCATATCTCCACGATATAGT 950
280 LphProTySerTyThrArgSerLysSerLysAspHisGluGluLeuS 297
951 GTTTCATATATCTATACAGCAAGTAAAGAAAGAAAGACAGAGAACTGT 1000
297 eLrLeuAlaIleSerGluAlaValArgAlaIleAspLysThrSerLysAsn 313
1001 CTCTAGTAGCCAGTGAAGCAGTTGCTGTATGAGAAACCTAGTAAAAAT 1050
314 ThrArgTyThrHisGlyHisGlySerGluThrLeuTyIrrLeuAlaProG1 330
1051 ACCAGTATACATATGCGCATGCTGCAAGAACTTATACCTAGCTCTG 1100
330 YGlyGlyAspAspThrIleTyTyAspLeuGlyIleLysTySerPhe... 345
1101 AGGTGGGAGCAGATGATGATGATGATTTGGCATCAAAATATTCGTTACAA 1150
346 .....Th 346
1151 TTGAACCTTCAGATACGGGCAATACGCAATTCCTGCTGCCGAGCCGTTAC 1200
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
1201 ATCAAAACCACCTGTAGAGAAAGCTTTGCCGCTGCTCTTAAAA 1243
seq_name: /cgn2.6/ptodata/1/ina/5A.COMB.seq:US-07-649-591B-2
seq_documentation_block:
; Sequence 2, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patln (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: hybridization probe
; LOCATION: 133 to 178

```

```

; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: potential clip site
; LOCATION: 380 to 382
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: signal sequence
; LOCATION: 41 to 106
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-07-649-591B-2

```

```

alignment_scores:
  Quality: 1832.00      Length: 415
  Ratio: 5.103          Gaps: 3
  Percent Similarity: 86.506  Percent Identity: 85.783

```

```

alignment_block:
  US-09-980-881-2 x US-07-649-591B-2 ..

```

```

Align seg 1/1 to: US-07-649-591B-2 from: 1 to: 1749

```

```

1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCysG1 17
  |||||
41 ATGAAGCTTTCAGCCCTTCAGCTTCCTGACCCCATGTTCTCTCTCTGCA 90
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
  |||||
91 GCACATGCTCTCGCGTTTCAGAGTGGCCAAAGTTCTACGCTCTCTCTTA 140
34 rGThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrGlu 50
  |||||
141 GAACCTCAGGCAAGTTCATGCTTACAGATCTTACTACAAACATATGAG 190
51 lIleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
  |||||
191 ATTGTCCTGCGACGCCGTAACAGCTGACCTTATTGGAAGAAAAACA 240
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
  |||||
241 AGTCATATTTTGTAAATGATCTGATGTCGACAAATGTGAAGCCCAT 290
84 euAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
  |||||
291 TAAATGTGAGGGAATTCATGATGCTGCTGCGCAGAGCTGGAAGAT 340
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
  |||||
341 CTATTCACACAGCAATTTCCACACACAGTCAAGCCCGGAGCCTCCGC 390
117 aSerTyrTyrGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 134
  |||||
391 ATCGTACTATGACAGTATCTACTCAATGAATCTATTCCTGGAGAG 440
134 lPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
  |||||
441 AATTATTAACGACAGAGCATCTGATATGCTTACAAAAATCCACATTGGA 490
151 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysG1 167
  |||||
491 TCCCATTTGAGAGTACCCACTCATGTGTTTAAAGTTTCGGAAGAAGA 540
167 uGlnThrAlaLysAsnAlaIleThrPheAspCysGlyLysHisAlaAspG 184
  |||||
541 ACAAAACGCAAAATGCCATATGATGAGCTGTGGAATTCATGCCAGAG 590
184 luTrpLysSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
  |||||
591 AATGATCTCTCGCTTCTGCTTGCTGTCATATGAGCCATATATACGTCAA 640

```

```

197 ..... 197
641 TTCTATGGAAATATAGGCAATATACAAATCTCTGAGGCTTGATGATT 690
197 ..... 197
691 CTATGTTATGCCGGTGTAAATGTGACGGTTATGACTACTCATGAGAAA 740
198 ..AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCys 213
  |||||
741 AGAATCGAATGTGAGAGAAAGAACCTTCTTCTATGCGCAACATCATTC 790
214 lIleGlyThrAspLeuAsnSerAsnPheValSerLysHisIleTrpCysGluG1 230
  |||||
791 ATCGGAACAGACCTGAAATAGCACTTTCCTCCAAACACTGGTGTGAGGA 840
230 uGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProG 247
  |||||
841 AGGTGATCTCACTCTCATGCTCGGAACCTGACTGTGAGACTTATCCTG 890
247 luSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
  |||||
891 AGTCAGAACCAAGACTGAAGGAGCGTACGTTCTTGAGAGAAATATATC 940
264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
  |||||
941 AACCATATTAAGCATATACATGATGATCATCTCCACACATATATAGT 990
280 lPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluAs 297
  |||||
991 GTTTCATATTCCTATACAGAAATGAAAGCAAAACCAACGTGAGACTGT 1040
297 erLeuValAlaSerGluAlaValAlaArgAlaIleAspLysThrSerLysAsn 313
  |||||
1041 CTTGATGAGCCAGTGAACAGCTGCTGCTATGAGAAACATGTAATAAT 1090
314 ThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProG1 330
  |||||
1091 ACCAGGTATACACATGGCCATGCTCAGAAACCTTATACCTGATGCTCG 1140
330 yGlyGlyAspAspTrpIleTyrAspLeuGlyLysTyrSerPhe... 345
  |||||
1141 AGGTGGGAGCATGTGATCTATGATTTGGGCATCAAAATTTGTTACAA 1190
346 .....Th 346
1191 TTGAATCTGAGATACGGGCAATACGATTTCTGCTGCGGAGCGTTAC 1240
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
  |||||
1241 ATCAAAACCCACCTGTAGAGAAAGCTTTGCGCCCTGCTCTAAAAA 1283

```

```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-277-540-2

```

```

seq_documentation_block:
; Sequence 2, Application us/08277540
; Patent No. 5474901

```

```

; GENERAL INFORMATION:
; APPLICANT: Dreyne, Dennis T., Eaton, Dan L.

```

```

; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESS: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd

```

```

; CITY: South San Francisco
; STATE: California

```

```

; COUNTRY: USA
; ZIP: 94080

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (genentech)

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-277-540-2

alignment_scores:
Quality: 1832.00      length: 415
Ratio: 5.103          Gaps: 3
Percent Similarity: 86.506      Percent Identity: 85.783

alignment_block:
US-09-980-881-2 x US-08-277-540-2 ..
Align seg 1/1 to: US-08-277-540-2 from: 1 to: 1749

1 MetLysLeuGysSerLeuAlaValLeuValProLLeuValLeuPheCysG1 17
|||||
41 ATGAAGCTTTGCAAGCTTCGACAGTCTGTACCCATGTTCTCTTCGTGA 90
|||||
17 uGlnHsValPheAlaPheGlnSerGlyGlnValLeuAlaLeuPro 34
|||||
91 GCAGCATGTTTGGCTTCAGAGTGGCCAAAGTTTACGCTCTCTCTA 140
|||||
34 rGHisSerArgGlnValGlnValLeuGlnAsnLeuThrThrTrpGlu 50
|||||
141 GAACCTCTAGGCAAGTTCAAGTCTACAGAACTTACTCAACATATGAG 190
|||||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
|||||
191 ATGTCTCTGCGCAGCGGTACAGCTGACCTTATGTAAGAAAAACA 240
|||||
67 nValHsPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
|||||
241 AGTCCATTTTGTGTAATGCATCTATGTCGACAAATGGAAGCCCAT 290
|||||
84 euAsnValSerGlyIleProCysSerValLeuValAspValGluAsp 100
|||||
291 TAAATGTGAGCGAATTCATGACAGTCTCTGCGCAGACGTGAAAGAT 340
|||||
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
341 CTTATATCAACACAGACTTTCCAAACGACACAGTCCAGCCCGAGCTCC 390
|||||
117 aSerTrpTrpGlnGlnIleSerLeuAsnGluIleTrpSerTrpIleG 134
|||||
391 ATCGTACTATGAACAGTATCACTCACTAAATGAAATCTTTTGGATNG 440
|||||
134 IupHeIleThrGluArgHisProAspMetLeuThrLysIleHisIleG 150

```

```

|||||
441 AATTATTAAGTGAAGGCAATCCTGATATGCTTACAAAAATCCATTGCA 490
|||||
151 SerSerPheGluLysTrpProLeuTrpValLeuLysValSerGlyLysG1 167
|||||
491 TCCCTCATTTGAGAAAGTACCCACTATGTTTAAAGTTCTTGAAAAAGA 540
|||||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisLArg 184
|||||
541 ACAAAACGCCAAAAATGCCATATGATGACTGTGGAAATCCATGCCAG 590
|||||
184 IupTrpIleSerProAlaPheCysLeuTrpPheIleGlnHis 197
|||||
591 AATGATCTCTCCGCTTTCGCTTGCTGCTTCATAGCCATATTAACTC 640
|||||
197 .....
641 TTCTATGGATTAATAGGCAATATACCAATCTCTAGCGCTTGATTT 690
|||||
197 .....
691 CTATGTTATCCGGTGTGTTAATGTGACGCTTATGACTACTATGAAAA 740
|||||
198 ..AsnArgMetTrpArgLysAsnArgSerPheTrpAlaAsnAsnHisCys 213
|||||
741 AGAATCAATGTGGAGAAAGAACCGTCTTCTATGCGAAACATCATTCG 790
|||||
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluG1 230
|||||
791 ATCGGAACAGACCTGAATAGAACTTGTCCAAACACAGTGTGACGA 840
|||||
230 uGlyAlaSerSerSerCysSerGluThrTrpCysGlyLeuTrpArg 247
|||||
841 AGGTGATCCAGTTCCTCATGCTCGGAACCTTGTGACTTTATCTG 890
|||||
247 IuSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
|||||
891 AGTCAAACCCAGAAAGTGAAGGCGTGTAGTTCTTGAAGAAATATAC 940
|||||
264 AsnGlnIleLysAlaTrpIleSerMetHisSerTrpSerGlnHisIle 280
|||||
941 AACCAATTAAGCAATACATCAACATGCAATTCATCTCCACCAATATG 990
|||||
280 IuPheProTrpSerTrpThrArgSerLysSerLysAspHisGluLys 297
|||||
991 GTTTCATATTCCTTATACAGAAATTAAGCAAAAGCATGAGAACTGT 1040
|||||
297 eTrpValAlaSerGluAlaValArgAlaIleAspLysHisSerLysAsn 313
|||||
1041 CTCTACTACCGAGTGAAGCAAGTTGCTGCTATGCAAAAACTATGTA 1090
|||||
314 ThrArgTrpThrHisGlyHisGlySerGluThrLeuTrpValProG1 330
|||||
1091 ACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTTACCTCG 1140
|||||
330 uGlyGlyAspAspTrpIleTrpAspLeuGlyIleLysTrpSerPhe... 345
|||||
1141 AGGTGGGAGCATGTGATCTATGATTTGGCATCAAAATATTCGTATACA 1190
|||||
346 .....
1191 TTGAAGCTTCGAGATACGGGACATACGGAATCTTGTCCGAGCGTTAC 1240
|||||
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
|||||
1241 ATCAAAACCCACTGTAGAGAAAGCTTTTCCCTGTCTCTATAAA 1283
|||||
seq_name: /cgn2_6/ptodata1/1na/5a_COMB.seq:US-08-430-787A-2
seq_documentation_block:
; Sequence 2, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:

```



```

? APPLICANT: Drayna, Dennis T., Eaton, Dan L.
? TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genentech, Inc.
? STREET: 460 Point San Bruno Blvd
? CITY: South San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94080
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patin (Genentech)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/430,787A
? FILING DATE: 27-APR-1995
? CLASSIFICATION: 514
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/277,540
? FILING DATE: 19-JUL-1994
? APPLICATION NUMBER: 08/167727
? FILING DATE: 15-DEC-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 07/95944
? FILING DATE: 14-OCT-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 07/649591
? FILING DATE: 01-FEB-91
? ATTORNEY/AGENT INFORMATION:
? NAME: Hasak, Janet E.
? REGISTRATION NUMBER: 28,616
? REFERENCE/DOCKET NUMBER: 689D1C1D1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415/225-1896
? TELEFAX: 415/952-9881
? TELEX: 910/371-7168
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1749 bases
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? US-08-430-787A-2

```

alignment_scores:

Quality: 1832.00	Length: 415
Ratio: 5.103	Gaps: 3
Percent Similarity: 86.506	Percent Identity: 85.783

alignment_block:

```

US-09-980-881-2 x US-08-430-787A-2
Align seg 1/1 to: US-08-430-787A-2 from: 1 to: 1749

```

1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCysG1 17
 41 ATGAAGCTTTCACGCTTCGACGCTGTCACCACTGTCCTCTCTGGA 90
 17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
 91 GCACGATGCTTCGCTTCAGAGTGCCAGATCTAGCTGCTCTCTA 140
 34 rGhrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyGlu 50
 141 GAACCTCTAGGCAAGTCAAGTCTTACAGAACTTACTACAACTATGAG 190
 51 lLeuValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
 191 ATTGTTCTCTGCGACCGGTACACGCTGACCTTATGTGGAAGAAAAACA 240

```

67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
  241 AGTCATTTTTCGTAATGCACTGATGTCGACAACTGAAAGCCCAT 290
  84 eUAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
  291 TAAATGTAGCGGAATTCATGCACTGCTGCTGCGACAGCTGGAAT 340
  101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
  341 CTTATTCAACAGCAGATTTCACAGACACAGTCACCCCGAGCCTCGC 390
  117 aSerTyTrpGluGlnIleTyHisSerLeuAsnGluIleTySerTrpIleG 134
  391 ATCGTACTATGACAGTATCTACTCACTAAATCAATCTTCTTGATAG 440
  134 lUpheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
  441 AATTTATAGTGAAGAGCATCTGATATGCTTACAAAATCCACATTCGA 490
  151 SerSerPheGluLysTyProLeuTyValLeuLysValSerGlyLysG1 167
  491 TCCTATTTGAGAACTACCCACTATGTTTAAAGTTTCGGAAGAA 540
  167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
  541 ACAAAACAGCAAAATGCCATATGATGATGATGATGATGATGATGATG 590
  184 lUTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
  591 AATGATCTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 640
  197 ..... 197
  641 TTCTATGGGATTAATGGGCAATATACATCTCCGAGGCTTGATGATT 690
  197 ..... 197
  691 CTATGTATGCCGGTGTAAATGTGAGCGGTATGACTACTATCATGAA 740
  198 ..AsnArgMetTrpArgLysAsnArgSerPheTyAlaAsnHisLys 213
  741 AGAATCGAATGTGAGAAAGAACCGTCTTCTTCTTCTTCTTCTTCTT 790
  214 lIleGlyThrAspLeuAsnSerAsnAspPheValSerLysHisIleTrpCysGluG1 230
  791 ATCGGACAGACCTGATAGAGACTTCTGCTTCCAAACACTGCTGAGAGA 840
  230 uGlyAlaSerSerSerSerCysSerGluThrTyTrpCysGlyLeuTyProG 247
  841 AGGTGCATTCAGATTCCTCATGCTCGGAAACCTACTGAGACTTATCTG 890
  247 lUserGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 263
  891 AGTCGAACCAAGAGTGAAGCAGTGTGCTAGTCTTGTAGAAAGAAATGC 940
  264 AsnGlnIleLysAlaTyrlleSerMetHisSerTySerGlnHisIleVa 280
  941 AACCGATTTAAAGCATATCATCATGATCATCATCTCCAGCAATATAGT 990
  280 lPheProTySerTyThrArgSerLysSerLysAspHisGluLeuS 297
  991 GTTTCATATTCCTTATACAGCAAGTAAAGCAAAACCATGAGAACTGT 1040
  297 eTrleuValAlaSerGluAlaValAlaArgAlaIleAspLysThrSerLysAsn 313
  1041 CTCTAGTAGCAGAGTGAAGCAGTTCGCTATTTGAAGAAACTAGTAAAT 1090
  314 ThrArgTyThrHisGlyHisGlySerGluThrLeuTyTrpLeuAlaProG1 330
  1091 ACCAGGTATACAGATGCGCATGCGCTCAGAAACCTATATACAGTCCCG 1140
  330 yGlyGlyAspAspTrpIleTyTrpAspLeuGlyIleLysTySer.Phe... 345

```

```

|||||
1141 AGGTGGGACGATTGATCTATGATTGGGCATCAAAATTCGTTACAA 1150
346 .....Th 346
1191 TTGAACCTCGATACGGGCACATACGATCTCTGCGCCGACCGTTAC 1240
346 rSerAsnProProValGluIysLeuProLeuSerLeuLys 360
|||||
1241 ATCAACCCACCTGTAGAGAGCTTTGGCCGCTGCTCTAATA 1283

```

seq_name: /cgn2_6/prodata/1/lna/5A_COMB.seq:us-08-696-139-1

seq_documentation_block:

```

; Sequence 1, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hersberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Steiner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1215 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1215
; US-08-696-139-1

```

alignment_scores:

```

Quality: 640.50 Length: 361
Ratio: 2.680 Gaps: 6
Percent Similarity: 66.205 Percent Identity: 37.673

```

alignment_block:

US-09-980-881-2 x US-08-696-139-1 ..

Align seq 1/1 to: US-08-696-139-1 from: 1 to: 1215

```

23 PheInSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnVa 39
|||||
28 TTCGAAGGGAGGAAGGTGTTCCGTCGCAATGTTCAAGATGAATGACAT 77
39 LgInValLeuGlnAsnLeuThrThrThrThrThrThrThrThrThrThr 56
|||||
78 CAGTCTACTCATCAGATGTTGGCCAGCACCAGCAGCAATGACTTGGGAAC 127
56 roValThrAlaAspLeuLeuValLysLysLysGlnValHisPhePheVal 72
|||||
128 CAGATCTCTCACACAATCAAACTCAGATCAGATGACTTCCGCTG 177
73 AsnAlaSerAspValAsnAsnValLysAlaHisLeuAsnValSerGly 89
|||||
178 AAAGCAGAGATATTTGGCTGGGAAGACTTTCTGGACAGCAATACAC 227
89 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 106
|||||
228 ACAATATAGAGTACTCATTAACAACCTGAGATCTGTCTGAGGCTCAGT 277
106 LeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGln 122
|||||
278 TTGACAGCAGAGTC.....CGTACAACCTGACACAGATATGAGAG 318
123 TyrHisSerLeuAsnGlnIleTyrSerThrIleGluPheLeuThrL 139
|||||
319 TACAACACCTGGGAAGACGATCGAGGCTTGACATCAACCAAGTCAGTA 368
139 gHisProPmetLeuThrLysIleHisIleGlySerSerPheGluLys 156
|||||
369 AAATCCAGACCTCATCTCCGACACACCATCCGACATACATTTTGA 418
156 yProLeuTyrValLeuLysValSerGlyLysGlnIleThrAlaLys 172
|||||
419 ACAATATATACCTCCCTCAAGGTT...GGCAACCTGACCAATATAG 465
173 AlaIleThrIleAspCysGlyIleHisAlaArgGluThrIleSerPro 189
|||||
466 GCCATTTTCATGACGACGTGTTCCATGCCAGAGATGATTTCCATGC 515
189 aPheCysLeuThrPheIle..... 195
|||||
516 ATTTCGCCAGTGTGTGTGAGAGAGGCTGTTCACACTATGATATGAGA 565
195 ..... 195
566 GTCACTATGACAGAAATCTCAACAAAGCTATGATTTATGCTTGCCCTG 615
196 .....GlyHis.....AsnArgMetThrPar 202
|||||
616 CTCGAATATGATGCGCTACATCTACACCTGACCAAGAACCGAATGTGAG 665
202 GlyAsnAlaGserPheTyrAlaAsnAsnHisCysIleGlyThrAspLeu 219
|||||
666 AAAGACCCGCTCTACCAATGCTGGAACCTACATGTCGACAGACCCCA 715
219 snSerAsnPheValSerLysHisTrpCysGluGluGluAlaSerSerSer 235
|||||
716 ACAGAAATTT...GATGCTGGGTGTGACAACTGAGACCTCTACAGAC 762
236 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGlu 252
|||||
763 CCTGCGATGACATTTACTGTGATCTGCTGACAGAGTCTGAAAAAGACAC 812
252 LlysAlaValAlaSerPheLeuLysArgAsnIleAsnGlnIleLysAla 269
|||||
813 CAAGCCCTGGCTGATTTATACCAACAACCTCTCCATCAAAAGCAT 862
269 yLleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 285
|||||
863 ACCTGACGATCCATCATCTACAGATGATATCTACCTTATTCCTAT 912

```

```

286 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG1 302
    ::::::::::::::::::::
913 GATTACAAACCCCGAGACAAATCGTAGTTGATAACCTGGCTAAAGC 962
    ::::::::::::::::::::
302 uAlaValAlaGluAlaIleAspLysThrSerLysAsnThrArgTyrThrHisG 319
    ::::::::::::::::::::
963 TGCCCTGAAACAACTT...GCTACACTGTATGCGACCAAGTACACATACG 1009
    ::::::::::::::::::::
319 LYNHISG1SerGluThrLeuTyrLeuAlaProGlyGlyAlaAspAspTrp 335
    ::::::::::::::::::::
1010 GCCCGAGGAGTACACAACTATCTGCTGCTGCTGGGGGCTGTGATGACTGG 1059
    ::::::::::::::::::::
336 IleTyrAspLeuGlyIleLysTyrSerPheThr 346
    ::::::::::::::::::::
1060 GCTTATGACCAAGAAATCAATATCTCTTACAC 1092
    ::::::::::::::::::::

```

seq.name: /cgn2_6/prodata/1/lna/5b_COMB.seq:US-08-860-882A-56

seq_documentation_block:

```

: Sequence 56, Application US/08860882A
: Patent No. 5985281
:
: GENERAL INFORMATION:
: APPLICANT: TAYLORSON, CHRISTOPHER JOHN
: APPLICANT: EGGLESE, HENDRIKUS JOHANNES
: APPLICANT: TARRAGONA-FIOL, ANTONIO
: APPLICANT: RABIN, BRIAN ROBERT
: APPLICANT: BOYLE, FRANCIS THOMAS
: APPLICANT: HENNAH, JOHN FREDERICK
: APPLICANT: BLAKELY, DAVID CHARLES
: APPLICANT: MARSHAM, PETER ROBERT
: APPLICANT: HEATON, DAVID WILLIAM
: APPLICANT: DAVIES, DAVID HOW
: TITLE OF INVENTION: CHEMICAL COMPOUNDS
: NUMBER OF SEQUENCES: 77
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PILLSBURY, MADISON & SUTRO
: STREET: 1100 NEW YORK AVENUE, N.W.
: CITY: WASHINGTON
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/860,882A
: FILING DATE: JUNE 23, 1997
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: DONALD J. BIRD
: REGISTRATION NUMBER: 25,323
: REFERENCE/DOCKET NUMBER: 9901/238653
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 861-3027
: TELEFAX: (202) 822-0944
: TELEX: 6174627 CUSH
: INFORMATION FOR SEQ ID NO: 56:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1263 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-860-882A-56

```

```

alignment_scores:
  Quality: 639.00      Length: 379
  Ratio: 2.526         Gaps: 7
  Percent Similarity: 66.755   Percent Identity: 35.620

```

alignment_block:
US-09-980-881-2 x US-08-860-882A-56 ..
Align seg 1/1 to: US-08-860-882A-56 from: 1 to: 1263

```

6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAl 22
  ::::::::::::::::::::
4 CTTCTTGTTCTGTGACTGACCTGGCCCTGGCANTGCTCATCAAGTGTGTA 53
  ::::::::::::::::::::
22 a...PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArg 38
  ::::::::::::::::::::
54 GCACCTTGAAGCGAGAAAGGTGTCTCGTTTACCTGTACCTGAAGTGAATC 103
  ::::::::::::::::::::
38 LlnValGlnValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrp 54
  ::::::::::::::::::::
104 ACATTAACATTAATCCGCGAGTGGCCAGACAGACCAGATTACCTTCGG 153
  ::::::::::::::::::::
55 GlnProValThrAlaAspLeuIleValLysLysGlnValHisPhePh 71
  ::::::::::::::::::::
154 AAGCCAGATTCTGTCACACAAATCAACCTCACAGTACAGTTGACTTCG 203
  ::::::::::::::::::::
71 eValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSer 88
  ::::::::::::::::::::
204 TGTTAAGCAGAGATACTGTCTGAGATGTTCTTAAGCAGATG 253
  ::::::::::::::::::::
88 LylLeProCysSerValLeuLeuAlaAspValGlnAspLeuLeuGln 104
  ::::::::::::::::::::
254 AACTCAATATACAGAGTACTGATATACCAACTGAAATGTGTGAGGCT 303
  ::::::::::::::::::::
105 GlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTrp 121
  ::::::::::::::::::::
304 CAGTTTGATGAGCCGGGT.....CTGCACAGGACACAGTTATGCA 344
  ::::::::::::::::::::
121 uGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThg 138
  ::::::::::::::::::::
345 GAAGTACACACAGGTGGAGACATAGAGCTTGACTACACAGTGGCCA 394
  ::::::::::::::::::::
138 LArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 154
  ::::::::::::::::::::
395 CTGAGATCCAGCCCTCATCTCTCGAGTGTATGGAACACATTTGAG 444
  ::::::::::::::::::::
155 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAla 171
  ::::::::::::::::::::
445 GGACGGCTATTACTCTCTGAAGTT...GGCAAGCTGGCAAAATTA 491
  ::::::::::::::::::::
171 sAsnAlaIleTrpIleAspCysGlyIleHisAlaArgIleTrpIleSerP 188
  ::::::::::::::::::::
492 GCCTGCCATTTTCATGCACTGTGTTCCATGCCAGAGATGATTTCTC 541
  ::::::::::::::::::::
188 roAlaPheCysLeuTrpPheIle..... 195
  ::::::::::::::::::::
542 CTGCATTTGCCAGGTGTTGTAAAGAGAGCTGTTGTACTATGAGAGCT 591
  ::::::::::::::::::::
195 ..... 195
  ::::::::::::::::::::
592 GAGATCCAAAGTACAGAGCTCTCGACAAGTTAGACTTTTATGCTCTGCC 641
  ::::::::::::::::::::
196 .....GLYHis.....AsnArgMet 201
  ::::::::::::::::::::
642 TGTGCTCAATATTGATGCTTACATCTACACCTGGACCAAGAGCCGATT 691
  ::::::::::::::::::::
201 rPArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 217
  ::::::::::::::::::::
692 GGAGAAAGACTGCGTCCACCACTAGCTGATGCTGATGCTGACAGAGC 741
  ::::::::::::::::::::
218 LeuAsnSerAsnPheValSerLysHisTrpCysGluGluAlaSerSe 234
  ::::::::::::::::::::
742 CCCAACGAAATTTT...GATGCTGGTGTGTGAATGAGACCTCTCG 788
  ::::::::::::::::::::
234 rSerSerCysSerGluThrTyrCysGlyLeuTyrProGlnSerGluProG 251
  ::::::::::::::::::::
789 AAACCCCTGTGATGAACATTACTGTGACCTGCCGAGAGTCTGAAGAG 838
  ::::::::::::::::::::

```

```

251 luvalysalavalaserpheluarlgarganlleasnginlleys 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
839 AGACCAAGCCGTGGATTTCATCGCAACAACCTCTTCATCAAG 888
268 AlATyrllesermechiserryserglnhisilevalpheprotyrse 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
889 GCATATGTGCATCCACTGTCTACTCCCAATGATGATCTTACTTACTC 938
284 rTyrrhrArgserlyserlyspahisglugluLeuserleuvalalas 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
939 ATATGCTTACAACTCGTGAGCAACATGCTGACTGAATGCCCTGCTGCTA 988
301 ergluvalavargalalleasplysrhrserlysaenhrargtyrrhr 317
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
989 AACCTACTGGAAGAACTT...GCCCTACTGCACGGCACCACCACTACACA 1035
318 HisglYHislysergluThrleuTyrrleuAlaproglYglYglYspas 334
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 TATGGCCCGGAGCTACACATCTATCCCTGCTGGGGGCTCTGACGA 1085
334 pTrrpIleTyrrAspleuGlylleYsrTyrrSerpherhr 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1086 CTGGCGCTTATGACCAGCAATCATATCTCTTCAAC 1122

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:us-08-860-882A-71

```

seq_documentation_block:
; Sequence 71, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860, 882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1284 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-860-882A-71

```

alignment_scores:

Quality:	631.50	Length:	361
Ratio:	2.588	Gaps:	6
Percent Similarity:	67.590	Percent Identity:	35.734

alignment_block:

US-09-980-881-2 x US-08-860-882A-71 ..

Align seg 1/1 to: US-08-860-882A-71 from: 1 to: 1284

```

23 PheGlnSerGlyGluValleuAlaAlaLeuProArgThrSerArgGlnVa 39
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 TTTCGAAGCCGAGAAAGTTCCTCCGTTACGTTACGATGAAATGACAT 134
39 lGlnValleuGlnAsnleuThrhrThrTyrrGluIlleValleuTrrpInr 56
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 TACATATATCCGCGAGTTGGCCACGACGACCAGATTGACTTCTGGAAAC 184
56 rovalThrAlaAspleuIlleVallylsysgluValhisPhepheVal 72
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 CAGATTCTGTACACAAATCAAACTCAGATGACTGACTTCCGCTT 224
73 AsnAlaSerAspValaAspAsnValLysAlaHisleuAsnValSerGlyI 89
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 AAAGCAGAAAGATACTGTCACTGTGAGAAATGTTTAAAGCAGAAATGAACT 284
89 eProCysSerValleuLeuAlaAspValaGluAspLeuIlleGlnGlnInr 106
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 ACAATACAAAGTACTGATTAAGCAACCTGAGAAATGTGTGAGAGCTAGT 334
106 leSerAsnAspThrValSerProArgAlaSerTyrrTyrrGluGln 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
335 TTGATGACCGCGTT.....CGTGCACAGACGACACACTTATGAGAAC 375
123 TyrrHisSerleuAsnGluIleTyrrSerTrpIleGluPheIleThrGluAr 139
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
376 TACAAACAAGTGGGAAACGATAGAGAGCTTGACATCAAAAGTCCCACTGA 425
139 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysr 156
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
426 GAATCCAGCCCTCATCTCTGCGAGTGTATCGGACACACATTTGAGAGGAC 475
156 yrProLeuTyrrValleuLysValSerGlyLysGluGlnThrAlaLysAsn 172
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 GCGCTATTATACCTCGTGAAGTT...GCCAAAGCTGGACAAATTAAGCC 522
173 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
523 GCCATTTCATGAGCTGTGTTCATGCCAGAGAGTGGATTTCTCTCGTC 572
189 aPheCysLeuTrpPheIle..... 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
573 ATTCTGCAAGTGTGTGTGAAGAGAGGCTGTCTACTATGACGAGTGA 622
195 ..... 195
623 TCCAACTGACAGAGCTTCTGACAAAGTTAGACTTTATGCTGCTGCTG 672
196 .....GlyHis.....AsnArgMetTrpAr 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
673 CTCATATATGATGGTTACATCTACACCTGAGACCAGAGCGGATTTGGAG 722
202 glyAsnArgSerPheTyrrAlaAsnAsnHisCysIleGlyThrAspLeuA 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
723 AAAGACTCCCTCACCACATACGTAGCTGATCTGATGGACAGACCCCA 772
219 snSerAsnPheValSerLysHisTrpCysGluGluValaIleSerSer 235
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
773 ACAGAAATTTT...GATCGTGTGGTGTGAAATGAGACCCCTCTGCAAAC 819
236 SerCysSerGluThrTyrrCysGlyLeuTyrrProGluSerGluProGluVa 252

```

```
|||||
820 CCCTGTGATGAACCTTACTGTGACCTGCCGACAGTCTGAAAAGAGAC 869
252 LlysAlaValAlaSerPheLeuArgAsnIleAsnGlnIleLysAla 269
|||||
870 CAAGCCCTGCGATTTTCATCCGCAACAACCTCTTCCATCAAGCAT 919
269 yTlieSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 285
|||||
920 ATCTGAACATCTACTGCTACTCCAAATGATGATCACTACCTTACTCATAT 969
286 ThrArgSerLysSerLysAspHisGlnGlnLeuSerLeuValAlaSerGln 302
|||||
970 GCTTACAAACCTGCGTGAACAATCTGATGTAATGCCCTGCTAAAGC 1019
302 uAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisG 319
|||||
1020 TACTGTGAANAACCTT..GCCTCACGTCACAGGACCAAGTACACATATG 1066
319 LysHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTyr 335
|||||
1067 GCCCGGAGGATCAACAATCATCTATCTGCTGCTGGGCGCTGACGACTGG 1116
336 IleTyrAspLeuGlyIleLysTyrSerPheThr 346
|||||
1117 GCTTATGACCAAGAAATCATATCTCTTGCAC 1149
seq_name: /cgn2_6/prodata/1/lna/5B_COMB.seq:us-09-171-945-124
seq_documentation_block:
; Sequence 124, Application US/09171945
; Patent No. 6277399
; GENERAL INFORMATION:
; APPLICANT: Emery, Stephen
; APPLICANT: Copley, Clive Graham
; APPLICANT: Edge, Michael Derek
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said
; FILE REFERENCE: Antibody, and their Therapeutic Use in an Adept System
; CURRENT FILING DATE: 1998-10-29
; PRIOR FILING DATE: 1997-04-29
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9703103.3
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 124
; LENGTH: 2154
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: humanized
US-09-171-945-124
alignment_scores:
Quality: 623.00 Length: 382
Ratio: 2.462 Gaps: 7
Percent Similarity: 66.230 Percent Identity: 34.555
alignment_block:
US-09-980-881-2 x US-09-171-945-124 ..
Align seg 1/1 to: US-09-171-945-124 from: 1 to: 2154
3 LeuGlySerLeuAlaValAlaLeuValProIleValLeuPheCysGlnGlnI 19
::: ||||| ||||| ||||| ||||| |||||
1 ATGTTGGACATCTTGGTCTGTGCTGACGTGGCCCTGGCATCTGCATCA 50
19 sValPheAla...PheGlnSerGlyGlnValLeuAlaIleAlaLeuProArgT 35
```

```
|||||
51 TGTGTGTAGACACCTTTGAAGCGAGACAGGTGTCCGTGTAACTTAAAG 100
35 hrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyrGluIle 51
|||||
101 ATGAANAATCACATTTATACATAATCCCGAGTTGGCCAGCACACCAAGTT 150
52 ValLeuTyrGlnProValThrAlaAspLeuIleValLysLysGlnVal 68
|||||
151 GACTCTGGAAGCCAGATTTCTGTACACAAATCAAACTCAACAGTACAGT 200
68 LHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisLeu 85
|||||
201 TGACTTCGCTGTTAAAGCAGAAATACATCTCACTGTGAGATGTTCTTAA 250
85 snValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 101
|||||
251 AGCAGATGAATACATACAAATACAGTACTGATTAAGCAACCTGGAATATG 300
102 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSe 118
|||||
301 GTGAGGCTCAGATTGATAGCGGGT.....CGTGAACAGGACA 341
118 rTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTyrPheGluP 135
|||||
342 CAGTTATAGAGAAATCAACAAGTGGGAAACGATACAGCTTGACTCTAAC 391
135 heIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySer 151
|||||
392 AAGTGGCCACGACAGAAATCCAGCCCTCATCTCTCGCAGTGTATTCGAAC 441
152 SerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluG 168
|||||
442 ACATTTGAGGAGCGCGCTATTACCTCCTGAAGTT...GGCAAGCTGG 488
168 nThrAlaLysAsnAlaIleTyrPheAspCysGlyIleHisAlaArgLut 185
|||||
489 ACAAATATAGCTCGCATTTTCATGACTGTGTTTCATGCGACAGAGT 538
185 rPheIleSerProAlaPheCysLeuTyrPheIle..... 195
|||||
539 GGATTTCTCTGCAATCTCTGCCAGTGTGTTGTAAGAGAGCTTGTGAC 588
195 ..... 195
589 TATGAGCGTGAAGTCAAGTACAGAGCTTCTGACAAAGTTAGACTTTTA 638
196 .....GlyHis.....A 198
639 TGTCTGCTGCTGTGTCATATATTGAGGCTACATCTACACTGACCAAGA 688
198 snArgMetTyrArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 214
|||||
689 GCCGATTTTGGAGAAAGACTGCTCCACCACATACGAGATCTAGCTCAT 738
215 GlyThrAspLeuAsnSerAsnPheValSerLysHisIstPyrCysGluGln 231
|||||
739 GGCACAGACCCCAAGAAATTTT...GATGCTGTGTGTGAATATGG 785
231 yAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGluS 248
|||||
786 AGCCTCTCGAACCCTCTGATGAACCTTACTGTGACCTGCCGACAGGT 835
248 ergLupProGluValLysAlaValAlaSerPheLeuArgAsnIleAsn 264
|||||
836 CTGAANAAGGACACCAAGCCCTGCTGATTTTCATCCGCAACAACCTCT 885
265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 281
|||||
886 TCCATCAAGGATATCTGACATCACTGCTACCTCCAAAGATGATGATA 935
281 eProTyrSerTyrThrArgSerLysSerLysAspHisGlnGlnLeuSerL 298
|||||
```

```

936 CCCCTACTCATATGCTTACAAACTCGGTGAGACAAATGCTGAGTTGAATG 985
298 euValAlaSerGluAlaValAlaArgAlaIleAspLysThrSerLysAsnThr 314
986 CCCGCTAAGACTACTGTGTAAGAACTT...GCCCTACGACGACGGCACC 1032
315 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGly1 331
1033 AAGTACACATATGCCCCGGAGGACTACAAACATATCTCTCTGCTGGGAC 1082
331 yGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThr 346
1083 TTCTAAGAACTGGGCTTATGACCAAGAAATCAGATATTCCTTCCACC 1128

seq_name: /cgn2.6/pldata/1/lna/5B.COMB.seq:US-08-782-760-5

seq_documentation_block:
; Sequence 5, Application US/08782760
; Patent No. 5948668
; GENERAL INFORMATION:
; APPLICANT: Hartman, Jacob
; APPLICANT: Fulga, Nelta
; APPLICANT: Mendelovitch, Simona
; APPLICANT: Gorecki, Marian
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,760
; FILING DATE: 13-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/778,233
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0336/43847
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 927 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..927
; US-08-782-760-5

alignment_scores:
Quality: 604.00 Length: 269
Ratio: 3.374 Gaps: 5
Percent Similarity: 66.543 Percent Identity: 44.610

```

```

alignment_block:
US-09-980-881-2 x US-08-782-760-5
Align seg 1/1 to: US-08-782-760-5 from: 1 to: 927

115 AlAserAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSe 131
1 GCAAGTTCACCAAGTTGCCACGTATATCCAGACTTGTCTCACAAGCG 50
131 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 148
51 GTGATTCACCAAGTTGCCACGTATATCCAGACTTGTCTCACAAGCG 100
148 lAlleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
101 TCATTTGAAACACATTTGAAAGAGCGTACATGTATGCTCCCAAGATT... 147
165 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 181
148 GGTAAACCTAGACCGAATAGCCCTGCCATCTTCATCGATTGTGTTCCA 197
181 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle..... 195
198 TGCACAGAGTGGATTTCCTCTCATTTCTGTCACTGTGTTGAGAGAG 247
195 ..... 195
248 CTGTCCGTACCTATATACCAAGAGATCCACATGAAACAGCTTCTAGTGA 297
196 ..... GlyHis..... 197
298 CTGGAATTTCTATGTTCTGCTGTGTCACAACTGATGATCTATGCTACAC 347
198 .....AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnA 211
348 CTGAGACTAAGAGACAGATGTGGAGAAAAACCCCTCTACTATAGCTGGGA 397
211 snHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrp 227
398 GTTCCTGCTTGGGTGTAAGACCCCAACAGAAATTT...AATGCTGCTGG 444
228 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 244
445 TGTGAAGTGGAGCTTCTCGAATCCCTGCTCTGGAACCTTACTGTGGACC 494
244 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 261
495 AGCCCAAGAGTCTGAAAAAGAGACAAAGCCCTGGCAGATTTCATCCGCA 544
261 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 277
545 ACAACTCTCCACATCAAGGCTACCTGACATCCATCACATCACTACACAG 594
278 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisI 294
595 ATGATGCTCTACCTTACTCTATGACTACAAACTGCTGGAACACTATGA 644
294 uGluLeuSerLeuValAlaSerGluAlaValAlaArgAlaIleAspLys 311
645 GGAATTTGAATGCCCTGTAAGGTGGCGCAAAAGAGCTT...GCCACTC 691
311 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 327
692 TGCATGGCAGCAAGTACATATATGCCCCGAGAGTCAACAACATCTATCT 741
328 AlaProGlyGlyLysAspAspTrpIleTyrAspLeuGlyIleLysTyrSe 344
742 GCTGCTGGGGGATGTGACGACTGTCTTATGATCAAGGAATCAATATATTC 791
344 rPheThr 346
792 CTTTACC 798

```


ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/696,139
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,258
FILING DATE: 16-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-8681
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..921
US-08-696-139-3

alignment_scores:
Quality: 576.00 Length: 264
Ratio: 3.310 Gaps: 5
Percent Similarity: 65.909 Percent Identity: 43.939

alignment_block:
US-09-980-881-2 x US-08-696-139-3 ..

Align seg 1/1 to: US-08-696-139-3 from: 1 to: 921

```
120 TyTgLuGlnIyHrHisSerLeuAnGluIleTySerThrIleGluPheI 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16 TATGAGAAAGTACAAACAAGTGGAAAGATGAGAGCTTGGACTAAGCAAGT 65
136 eThrgLuAgnHisProAspMetLeuThrIlyHisIleGlySerSerp 153
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 CACCAGTGAATAATCCAGACCTCATCTCGCACAGCCATCGAAGTACAT 115
153 heGluIyStyPrLeuIyValLeuIyValSerGlyIySgluGlnThr 169
|||:|||||:|||||:|||||:|||||:|||||:|||||:
116 TTTTAAAGAAACAATATATACCTCTCAAGGTT...GGCAAACTGAGACA 162
170 AAlLyAsnAlaIleTrrIleAspCysGlyIleHisAlaIarGluTrpI 186
|||:|||||:|||||:|||||:|||||:|||||:|||||:
163 AATAAGCGCTGCCATTTTCATGAGCTGTGTTCCATGCCAGAGAAATGAT 212
186 eSerProAlaPheCysLeuThrPheIle..... 195
|||||:|||||:|||||:|||||:|||||:|||||:
213 TTCCCATGCAATTTTGGCAGTGTGTGTGAGAGAGCGTGTTCACCTATG 262
195 ..... 195
263 GATATGAGAGTCAATGACAGAAATTCCTCAACAAGCTTAGACTTTATGTC 312
196 .....GlyHis.....AsnAr 199
313 TTGGCGTGTCTCAATATGATGGCTACATCTACACTGAGCAAGAACCG 362
199 gHeTrrAtrgLyAsnArSerPheTherIAlaAsnAsnHisCysIleGlyT 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
363 AATGTGAGAAAGACCCCTCTACCAATGTGTGAACTACCTGCATTGCCA 412
216 hrAspLeuAsnSerAspPheValSerIyHisTrpCysGluGlnAla 232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
413 CAGACCCCAAGAAATTTT...GATGCTGGGTGGTGCACAACCTGGAGCC 459
233 SerSerSerSerCysSerGluThrTyrcysGlyLeuTyrrProGluSerG 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 TCTACAGACCCCTCGATGAGACTTACTGTGGATCTCGTGCAGAGTGTGA 509
249 uProGluValIyAlaValAlaSerPheLeuArgArgAsnIleAsnGlnI 266
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510 AAAAGAGACCAAGGCGCTGGCTGATTTATAGCAACAACCTCTCCCA 559
266 lelysAlaTyrlleSerMetHisSerTyrrSerGlnHisIleValPhePro 282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 TCAAGCATACCTGACATCAGTACATACACAGATGATACCTTACCT 609
283 TySerTyrrThrArgSerIySerIySerIyAspHisGluGluLeuSerIe 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
610 TATTCCTATGATTACAAACTCCCGAGACAATGCTGAGTTGAATAACT 659
299 lAlaSerGluAlaValArgAlaIleAspIyThrSerIyAsnThrArgT 316
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
660 GGCTAAGGCTCGCTGAAGAACTT...GCTACACTGTATGACCAAGT 706
316 yTrHisGlyHisGlySerGluThrLeuTyrrLeuAlaProGlyGlyGly 332
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 ACAATATGCGGCCGAGAGCTACCAACAATCATCTGCTGCTGGGGGCTCT 756
333 AspAspTrpIleTyrrAspIleGlyIleIySerPheThr 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
757 GATGACTGGGCTTATGACCAAGGAATCAATATCTCTCAC 798
```

seq_name: /cgn2_6/plodata/1/lna/5B_COMB.seq:US-08-860-882A-67

seq_documentation_block:
Sequence 67, Application US/08860882A
Patent No. 5985281
GENERAL INFORMATION:
APPLICANT: TAYLORSON, CHRISTOPHER JOHN
APPLICANT: EGGELTE, HENDRIKUS JOHANNES
APPLICANT: TARRAGONA-FIOL, ANTONIO
APPLICANT: RABIN, BRIAN ROBERT
APPLICANT: BOYLE, FRANCIS THOMAS
APPLICANT: HENNEM, JOHN FREDERICK
APPLICANT: BLAKELY, DAVID CHARLES
APPLICANT: MARSHAM, PETER ROBERT
APPLICANT: HEATON, DAVID WILLIAM
TITLE OF INVENTION: CHEMICAL COMPOUNDS
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:

Alignment block:
 US-09-980-881-2 x US-08-860-882A-64 ..
 Align seg 1/1 to: US-08-860-882A-64 from: 1 to: 1053

```

115 A1aSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSe 131
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 GCAACTGGTCACCTTTACGAGAGTACAAACAGTGGGAAACGATAGAGCC 116
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
131 rTrrilegUpheileThrGluArgHisProAspMetLeuThrLysIleH 148
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 TTGGACTCAACAAGTCGCCACTGAGAAATCCAGCCCTCATCTCCGCGAGT 166
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
148 IsilegIleSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 TTAATCGAACAACATTGAGGAGCGGCTATTACCTCTGAAAGTT... 213
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
165 G1LYSG1G1nThralLysAsnAlaIleTrrPleAspCysG1LYLeH1 181
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
214 GCGMAAGCTGCGACAAATAGAGCTGCCATTTTCATGAGCTGGTTCCCA 263
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 sAlaArgGluTrrPleSerProAlaPheCysLeuTrrPhele..... 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
264 TGCACAGAGAGTGATTTCTCTGCATTTCTGCCAGTGTGTTGTAAGAGG 313
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
195 ..... 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
314 CTGTTCGTACSTATGAGAGCTGAGATCCAAAGTACAGAGCTTCTGCAAG 363
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
196 ..... 197
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 TTAGACTTTATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 413
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
198 ..... 211
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
414 CTGACCAAGAGCCGATTTTGGAGAAAGAGCTGCTGCCACCCATCTGGAT 463
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
211 sHnHsCysAlleG1LyrHisPheLeuAsnSerAsnPheValSerLysH1sTrr 227
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
464 CTAGCTGCTTGGACAGAGCCGACAGCCCAAGAAATTT...GATCTGCTTGG 510
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
228 CysG1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1 244
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
511 TGTGAATGAGAGCCCTGCGAAACCCCTGTGATGAACCTTACTGTGAGCC 560
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
244 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 261
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
561 TGCSCCAGAGTCTGAAGAGAGAGCAAGCCCTGCTGATTCATCCGCA 610
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
261 rGAsnIleAsnGlnIleLysAlaTyrLieserMetHisSerTyrSerGln 277
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
611 ACMAACTCTCTTCATCAAGGATATGACAAATCCATCCTGATCCCA 660
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
278 HisIleValPheProTyrSerTyrTrrArgSerLysSerLysAspHisG1 294
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
661 ATGATGATCTACCTTACTCATATGCTTACAAACGCGGAGAACATATC 710
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
294 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleAspLysTrs 311
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
711 TGAGTTGAATGCCCTGCTAAAGCTAGCTGAAGAACTT...GCCTGAC 757
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
311 eTlysAsnThrArgTyrTrrHisGluHisGlySerGluThrLeuTyrLeu 327
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
758 TGCACGGAGCAAGTACATATGCGCCGCGGAGTACAAAGATATATCTCT 807
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
328 AlaProG1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1 344
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
808 GCTGGTGGGGGCTGAGAGCTGGGCTTATGACCAAGATCAGATATTC 857
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
344 rPheThr 346
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
858 CTTCACCC 864
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-860-882A-74
 seq_documentation block:
 ; Sequence 74, Application US/08860882A
 ; Patent No. 5985281
 ; GENERAL INFORMATION:
 ; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
 ; APPLICANT: EGGELE, HENDRIKUS JOHANNES
 ; APPLICANT: TARRAGONA-FIOL, ANTONIO
 ; APPLICANT: RABIN, BRIAN ROBERT
 ; APPLICANT: BOYLE, FRANCIS THOMAS
 ; APPLICANT: HENNAM, JOHN FREDERICK
 ; APPLICANT: BLAKELY, DAVID CHARLES
 ; APPLICANT: MARSHAM, PETER ROBERT
 ; APPLICANT: HEATON, DAVID WILLIAM
 ; APPLICANT: DAVIES, DAVID HUM
 ; TITLE OF INVENTION: CHEMICAL COMPOUNDS
 ; NUMBER OF SEQUENCES: 77
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PILLSBURY, MADISON & SUTRO
 ; STREET: 1100 NEW YORK AVENUE, N.W.
 ; CITY: WASHINGTON
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy Disk
 ; COMPUTER: IBM compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/860, 882A
 ; FILING DATE: JUNE 23, 1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: DONALD J. BIRD
 ; REGISTRATION NUMBER: 25,323
 ; REFERENCE/DOCKET NUMBER: 9901/238653
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 861-3027
 ; TELEFAX: (202) 822-0944
 ; TELEX: 6174627 CUSH
 ; INFORMATION FOR SEQ ID NO: 74:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1059 bases
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-860-882A-74

alignment_scores:
 Quality: 554.00 Length: 269
 Ratio: 3.095 Gaps: 5
 Percent Similarity: 66.543 Percent Identity: 40.520

Alignment block:
 US-09-980-881-2 x US-08-860-882A-74 ..
 Align seg 1/1 to: US-08-860-882A-74 from: 1 to: 1059

```

115 A1aSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSe 131
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 GCAACTGGTCACCTTTACGAGAGTACAAACAGTGGGAAACGATAGAGCC 116
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
131 rTrrilegUpheileThrGluArgHisProAspMetLeuThrLysIleH 148
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 TTGGACTCAACAAGTCGCCACTGAGAAATCCAGCCCTCATCTCCGCGAGT 166
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
148 IsilegIleSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 TTAATCGAACAACATTGAGGAGCGGCTATTACCTCTGAAAGTT... 213
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

165 GlyylsGluGlnThrAlaLysAsnAlaIleTrrPleaspGlyIleH 181
166 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 GGCAGAACTGGACAAATAAGCCCTGCTCAATGAGCTGTGGTTCCA 263
168 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 sAlaArgGluTrrPleSerProAlaPheCysLeuTrrPhele..... 195
170 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
171 TGCCAGAGATGGATTCTCTGCTGATCTGCACTGGTTTGAAGAGAG 313
172 ..... 195
173 314 CTGTTCTACCTATGAGCTGAGATCCAGTACAGAGACTTCTGCAGAG 363
174 ..... 197
175 196 ..... 197
176 364 TTAGACTTTTATGCTGCTGCTGCTCAATGATGCTTACATCTAC 413
177 ..... 211
178 198 ..... 211
179 414 CTGACCAAGAGCCGATTGAGAAAGACCTCCACCCATCTGAT 463
180 ..... 227
181 211 snHscysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrr 227
182 ..... 510
183 464 CTACCTGATGCGACAGACCCCAAGAAATTT...GATGCTGGTTGG 510
184 ..... 560
185 228 CysGluGluGlyAlaSerSerSerCysSerGluTrrTyrCysGlyLe 244
186 ..... 610
187 511 TGTAATTTGGAGCTCTCGAACCCTGATGATGAAATCTTGAGACC 560
188 ..... 610
189 244 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuA 261
190 ..... 610
191 561 TGCCGAGAGCTGTAAGAGAGACCAAGCCCTGCTGATTTCAATCCCA 610
192 ..... 277
193 261 rGAsnIleAsnGlnIleLysAlaTrrLysSerMetHisSerTyrSerGln 277
194 ..... 660
195 611 ACAAGCTCTCTCATGACGATATGACAAATCCATCTGATCCCA 660
196 ..... 294
197 278 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 294
198 ..... 710
199 661 ATGATGATCTACCTCTACTATGCTTACCAACCTCGGTGAGAACAAAG 710
200 ..... 311
201 294 uGluLeuSerLeuValAlaSerGluAlaValAlaIleAspLysTrrs 311
202 ..... 757
203 711 TGAATTTGATGCTGCTGCTAAGCTACTGTGAAGAACTT...GCCTCAG 757
204 ..... 807
205 758 TGACAGGAGACCAAGTACATATGAGCCCGGAGCTACAAACATATACCT 807
206 ..... 344
207 328 AlaProGlyGlyGlyAspAspTrrPleTyrAspLeuGlyIleLysTyrSe 344
208 ..... 857
209 808 GCTGCTGGGGCTCTAAAGACTGGCTTATGACCAAGAAATCATGATATTC 857
210 ..... 346
211 344 rPheThr 346
212 ..... 864
213 858 CTTCACC 864

```

seq_name: /c9n2_6/prodata/1/lna/5B_COMB.seq:US-08-860-882A-77

seq_documentation_block:

Sequence 77, Application US/08860882A

Patent No. 5985281

GENERAL INFORMATION:

APPLICANT: TAYLORSON, CHRISTOPHER JOHN

APPLICANT: EGGEITE, HENDRIKUS JOHANNES

APPLICANT: TARRAGONA-FIOL, ANTONIO

APPLICANT: RABIN, BRIAN ROBERT

APPLICANT: BOYLE, FRANCIS THOMAS

APPLICANT: HENNAM, JOHN FREDERICK

APPLICANT: BLAKELEY, DAVID CHARLES

APPLICANT: MARSHAM, PETER ROBERT

APPLICANT: HEATON, DAVID WILLIAM

APPLICANT: DAVIES, DAVID HENRY

TITLE OF INVENTION: CHEMICAL COMPOUNDS

```

NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESS: PILLSBURY, MADISON & SUTRO
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 1059 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-860-882A-77

alignment_scores:
Quality: 553.00 Length: 269
Ratio: 3.089 Gaps: 5
Percent Similarity: 66.543 Percent Identity: 40.520

alignment_block:
US-09-980-881-2 x US-08-860-882A-77 ..
Align seg 1/1 to: US-08-860-882A-77 from: 1 to: 1059

115 AlaSerAlaSerTrrTrrGluGlnTrrHisSerLeuAsnGluIleLysrSe 131
116 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 GCAACTGTCACACTTACAGAGAGTACAAAGTGGAGAAACATAGAGGC 116
118 ..... 148
119 131 tTrrPleGluPheIleThrGluArgHisProAspMetLeuTrrLysIleH 148
120 ..... 166
121 117 TGGACTCAACAAGTCGCCATGAGAAATCCAGCCCTCATCTCTCGCAGTG 166
122 ..... 164
123 148 tIleGlySerSerPheGluLysTrrProLeuTrrValLeuLysValSer 164
124 ..... 213
125 167 TTAATCGAAGCAACATTTGAGGAGCGGCTATTTACCTCTGAAAGCTT... 213
126 ..... 181
127 165 GlyLysGluGlnThrAlaLysAsnAlaIleTrrPleaspGlyIleH 181
128 ..... 263
129 214 GGCAGAACTGGACAAATAAGCCCTGCTCAATGAGCTGTGGTTCCA 263
130 ..... 195
131 181 sAlaArgGluTrrPleSerProAlaPheCysLeuTrrPhele..... 195
132 ..... 313
133 264 TGCCAGAGATGGATTCTCTGCTGATCTGCACTGGTTTGAAGAGAG 313
134 ..... 195
135 195 ..... 195
136 314 CTGTTCTACCTATGAGCTGAGATCCAGTACAGAGACTTCTGCAGAG 363
137 ..... 197
138 196 ..... 197
139 364 TTAGACTTTTATGCTGCTGCTGCTCAATGATGATGCTTACATCTAC 413
140 ..... 211
141 198 ..... 211

```

```
414 CTGACCAAGACCGATTGGAAGACCTCCACCATACGAT 463
      ::::::::::::::::::::
211 snH1sCys11eglyThrAspLeuAsnSerAsnPheValSerLysHisTrp 227
      ::::::::::::::::::::
464 CTAGCTGCATTGGCACAGACCCCAAGAAATTTT...CATGCTGTTGG 510
      ::::::::::::::::::::
228 CysgluglyAlaSerSerSerCysSerLysThrCysGlyLe 244
      ::::::::::::::::::::
511 TGTGAAATTTGAGCCTCTCGAAGCCCTGTGATGAACTTACTGTGACC 560
      ::::::::::::::::::::
244 UTyrProGluSerGluProGluValLysAlaLaserPheLeuArg 261
      ::::::::::::::::::::
561 TGCCCGCAGAGTCTGAAAGAGAGACCAAGGCCCTGGCTGATTTCATCCGA 610
      ::::::::::::::::::::
261 rGAsn11eAnG11eLysAlaTyrLieserMethisSerLysSerGln 277
      ::::::::::::::::::::
611 ACAAACTCTCTTCATCAAGGCATATCTGACAATCCACTGTACTCCAA 660
      ::::::::::::::::::::
278 His11eValPheProLysSerLysThrArgSerLysSerLysAspHisG1 294
      ::::::::::::::::::::
661 ATGATGATCTTACCTTACTCATATGCTTACAAACTCGTAGACAAATGC 710
      ::::::::::::::::::::
294 uG11eUeSerLeuValAlaSerGluAlaValArgAla11eAspLysTrs 311
      ::::::::::::::::::::
711 TGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAGAACTT...GCCCTCAC 757
      ::::::::::::::::::::
311 eRlyAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 327
      ::::::::::::::::::::
758 TGCACGGCACAAGTACACATATGGCCCGGAGCTACACAACTATCTTCCT 807
      ::::::::::::::::::::
328 AlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly11eLysTrise 344
      ::::::::::::::::::::
808 GCTGCTGGGGGCTCTCGGACTGGGCTTATGACCAAGAAATCAGATATTC 857
      ::::::::::::::::::::
344 rPheThr 346
      ::::::::::::::::::::
858 CTTCCACC 864
```

OM of: US-09-980-881-2 to: EST:* out_format: pfs

Date: Sep 18, 2002 5:41 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+gzn_model -DEV=xlh
-o=/cgn2.1/USP0_spool/US09980881/runat_16092002_140128_10623/app-query.fasta_1.1126
-DB=EST -OPMT=fastap -SUFFIX=first -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPEXT=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -TGAPOP=10.000 -TGAPEXT=0.500 -DELOPE=6.000
-DELETE=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODAL=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09980881@cgnl.1.3434
-MCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLIPX -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-980-881-2
Query length: 360
Database: EST:*
Database sequences: 13736207
Database length: 1841457050
Search time (sec): 3115.820000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb.htc:AK004045	1534.50	3066.48	1.5e-161	1446	AK004045 Mus musculus 18 days
gb.est1:AL536069	1324.50	2648.24	2.9e-138	889	AL536069 AL536069 lrtl_F1013.Pft
gb.est2:BG616629	1034.00	2063.27	1.1e-105	788	BG616629 602646064F1 NIH_MGC_76
gb.est2:BT144644	1015.50	2025.01	1.5e-103	853	BT144644 602909932F1 NCI_CGAP_1
gb.est2:BG616456	996.50	1968.94	1.7e-101	738	BG616456 602614618F1 NIH_MGC_76
gb.est2:BG656051	983.00	1960.95	5.1e-100	747	BG656051 602584562F1 NIH_MGC_76
gb.est2:BG618239	975.00	1945.15	4.2e-99	725	BG618239 602646567F1 NIH_MGC_76
gb.est2:BG6569281	965.00	1923.10	7.2e-98	852	BG6569281 602588573F1 NIH_MGC_76
gb.est2:BT1332659	957.50	1905.85	6.5e-97	1022	BT1332659 602884088F1 NCI_CGAP_1
gb.est1:AV646979	949.00	1892.51	3.6e-96	735	AV646979 AV646979 GIC Homo sapi
gb.est2:BG618813	927.50	1848.88	9.8e-94	750	BG618813 602646166F1 NIH_MGC_76
gb.est2:BG656100	902.00	1797.45	7.1e-91	747	BG656100 602583700F1 NIH_MGC_76
gb.est2:BG656240	898.00	1790.28	1.8e-90	691	BG656240 602583700F1 NIH_MGC_76
gb.est1:AI048167	884.50	1762.49	6.3e-89	724	AI048167 ud71812.Y1 Sugano mus
gb.est1:AV693037	868.00	1731.69	3.3e-87	583	AV693037 AV693037 GIC Homo sapi
gb.est1:AV658390	857.00	1709.12	5.9e-86	602	AV658390 AV658390 GIC Homo sapi
gb.est1:AV694058	849.00	1692.99	4.7e-85	601	AV694058 AV694058 GIC Homo sapi
gb.est1:AV698425	838.00	1670.41	8.5e-84	621	AV698425 AV698425 GIC Homo sapi
gb.est2:BG617911	837.50	1667.23	1.3e-83	589	BG617911 602645566F1 NIH_MGC_76
gb.est1:AV651709	815.50	1629.87	1.5e-81	688	AV651709 AV651709 GIC Homo sapi
gb.est2:BG6567264	808.50	1606.11	3.3e-80	806	BG6567264 602589455F1 NIH_MGC_76
gb.est1:AL532872	806.00	1606.11	3.3e-80	661	AL532872 u183811.Y1 Sugano mus
gb.est1:AL182113	806.00	1606.43	3.1e-80	580	AL182113 ud73110.Y1 Sugano mus
gb.est1:AA968307	806.00	1606.43	3.1e-80	589	AA968307 uc71107.Y1 Sugano mus
gb.est1:AM950700	803.50	1601.11	6.2e-80	603	AM950700 EST362270 MAGE resequ
gb.est1:AM950668	799.50	1593.33	1.6e-79	575	AM950668 AL536068 lrtl_F1013.Pft
gb.est2:N98450	773.50	1541.90	1.2e-76	537	N98450 za28009.X1 Soares fetal
gb.est1:AL532876	767.00	1527.09	1.6e-75	958	AL532876 mj24006.Y1 Soares mus
gb.est1:AL574117	766.00	1527.34	7.9e-76	511	AL574117 u196809.Y1 Sugano mus
gb.est1:AL528146	759.00	1510.21	4.9e-75	511	AL528146 u194804.Y1 Sugano mus
gb.est1:AV681868	742.00	1480.06	3.4e-73	462	AV681868 AV681868 GIC Homo sapi
gb.est2:BG216526	741.50	1472.10	9.5e-73	843	BG216526 RST36106 Atherys RAG
gb.est2:W88434	731.00	1448.03	5.8e-72	455	W88434 zhs9f04.X1 Soares fetal
gb.est1:AV658178	723.50	1449.35	1.8e-71	623	AV658178 AV658178 GIC Homo sapi
gb.est1:AV658128	718.00	1428.31	2.4e-70	615	AV658128 AV658128 GIC Homo sapi
gb.est1:AL574064	707.00	1406.20	4.4e-69	610	AL574064 u167911.Y1 Sugano mus
gb.est1:BE235838	694.00	1381.28	1.1e-67	447	BE235838 143448 MARC lptg Sus s
gb.est1:AL574057	688.00	1370.45	4.3e-67	584	AL574057 u167904.Y1 Sugano mus
gb.est1:AM413090	685.50	1360.84	1.5e-66	723	AM413090 uq52c03.X1 Sugano mus

gb.est1:AL525929 - 672.50 1333.84 4.8e-65 772 | AL525929 u183811.X1 Sugano m
gb.est2:BE384322 + 663.50 1316.94 4.2e-64 692 | BE384322 602046670F1 NCI_CGA
gb.htc:AK003061 + 661.00 1303.51 2.3e-63 1429 | AK003061 Mus musculus adult
gb.est1:AB050299 - 651.50 1296.02 6.1e-63 520 | AB050299 AB050299 lambda zap
gb.est1:AL525397 - 646.50 1281.74 3.8e-62 747 | AL525397 u194804.X1 Sugano m

seq_name: gb.htc:AK004045

seq_documentation_block:

LOCUS AK004045 1446 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus 18 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:1110032P04:carboxypeptidase B2 (plasma),
full insert sequence.

ACCESSION AK004045
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,
clone:1110032P04.
MUS musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

JOURNAL

MEDLINE 99279253
PubMed 10349636

REFERENCE

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

MEDLINE 20499374
PubMed 11042159

REFERENCE

Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamanoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohtsuka, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, D.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multiplexillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

MEDLINE 20530913
PubMed 11076861

REFERENCE

The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL

MEDLINE 5 (bases 1 to 1446)

Aach, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arawaka, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numata, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Schmitt, L., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Sohrishi, H., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tasunashi, A., Tejima, T., Toyota, T., Yamada, T., Yamana, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.

TITLE

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

189 GTCCTTCCTCGGACGACAGTACAGCTGAAATTCCTCCAGAAAGAA 238
188 GTCCCTTCCTCGGACGACAGTACAGCTGAAATTCCTCCAGAAAGAA 239
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysLahISL 84
239 AGTCACATTTTTTTGTAATGCGCTGATGTCGACAGTGCACAAACGCAAT 288
84 euAsnValSerGlyIleProCysSerValLeuLeuAlaAspValISLasp 100
289 TAAATGTGACGAGAAATTCATTTAAAGTCTGTGATGAACAAACGTGAGAC 338
101 LeuIleGlnGlnIleIleSerAsnAspThrValSerProArgAlaSerAl 117
339 CTAAATGAAACGACAGCACTTTCAAATGAAACGCTGACCCCGCGCTCCG 388
117 aSerTyrTyrGlnIleIleIleHisSerIleuAsnGlnIleTyrSerTrpIleG 134
389 TTCACTACTATGAGAGAGATACCTCGTCAATTAATGAATATATCTCGATAG 438
134 LurPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
439 AAGCATTAACGTGAACACACATCTGCATCTGCCAGAAATTCATCATGGA 488
151 SerSerPheGlnLysTyrProIleuTyrValLeuLysValSerGlyLysG 167
489 TCATCATTCGACGAAGATACCCACATTTATGTTTAAAGTCTCCAGAAAGGA 538
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgS 184
539 ACAAGAAATCAAAAAATGCCATCTGTGATCGACTGCGAATCCATGCGCAG 588
184 LurTrpIleSerProAlaPheCysLeuThrPheIleGlnHis 197
589 AATGATTTCACTGCTTCTTGTGTGGTTCATTAAGGTCACGTGACCAAA 638
197 197
639 TTCATGCGGAAAGAAATATGTTACCAAGACTTCTGAGGCACACTGATTT 688
197 197
689 CTACATCATGCCCGTGATGACAGCTGATGGCTATGACTACAGCTGAGAAA 738
198 . AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCys 213
739 AGAATGAGATGTGGAGAGAAACCGCTTCTGCTCACAAGAACAAACCGCTGC 788
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGlnI 230
789 GTGGGCACAGACTGAAACAGAACTTCGCTCCAAACACTGGTGAGAGA 838
230 uGlnAlaIleSerSerSerCysSerGlnIleThrTyrCysGlyLeuTyrProG 247
839 AGGTGGGTAAAGTCTCTCTGCTCTGAAACCACTGTGACCTTTATCTGTG 888
247 LurSerGlnProGlnValLysAlaValAlaIleSerPheLeuArgArgAsnIle 263
889 AGCTGAGCAGCAGGAGTGAAGGCGATGGCTTCTTGAGAAAGAAATATC 938
264 AsnGlnIleLysAlaIleTyrLleSerMetHisSerTyrSerGlnHisIleVa 280
939 GACCAACATTTAAAGCTTACATCAAGTATGCACTCATGCTCCCAACAAATACT 988
280 LurPheProTyrSerTyrThrArgSerLysSerLysAsnHisGlnIleuAs 297
989 GTTTCCGTAATCTCTTAACACAGAACAAAACGACAGACCAACGAAGACTGT 1038
297 eTyrLeuValAlaSerGlnAlaValArgAlaIleAspLysThrSerLysAsn 313
1039 CTCGTAGTGCGCCAGCAACAGATTTCTGTGAATTTGAAGATTAATTAANAAC 1088
314 ThrArgTyrThrHisGlyHisIleSerGlnIleThrLeuTyrLeuAlaProG 330

```

1089 ACCAGTACACACAGGCGAGTGGCTCAGAAAGTTATATCTAGCTCTCG 1138
330 YGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer.Phe... 345
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1139 AGTTCTGACGATGATGATCTATGATTTGGCATCAATATATCGTTTACA 1188
346 .....Th 346
1189 TTGAGCTCCGAGATACAGGAGATACGATTTCTGCGCTGAGAGATAC 1238
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1239 ATCAACCCACTTGTGCGAGAAAGCTTTGGCCCGCATCTCTAATA 1281

seq_name: gb_est1:AL536069

seq_documentation_block:
LOCUS AL536069 889 bp mRNA linear EST 13-FEB-2001
DEFINITION AL536069 ltr_FL013_FBrn1 Homo sapiens cDNA clone CS0DF022YE21 5
prime, mRNA sequence.
ACCESSION AL536069
VERSION AL536069
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 889)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..889
location/Qualifiers
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone_lib="CS0DF022YE21"
  /dev_stage="PFI_FL013_FBrn1"
  /lab_host="DH10B"
  /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
  cDNA was primed with a NotI-oligo(dT) primer. Five prime
  end enriched, double-stranded cDNA was digested with Not I
  and cloned into the Not I and Eco RV sites of the
  pCMVSPORT 6 vector. Library was constructed by Life
  Technologies. Contact : Feng Liang Life Technologies, a
  division of Invitrogen 9800 Medical Center Drive Rockville
  , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
  fliang@lifetech.com URL :
  http://fulllength.invitrogen.com"
BASE COUNT 255 a 186 c 192 g 235 t 1 others
ORIGIN

alignment_scores:
  Quality: 1324.50 Length: 294
  Ratio: 5.174 Gaps: 1
Percent Similarity: 87.075 Percent Identity: 86.395

alignment_block:
US-09-980-881-2 x AL536069 ..
Align seg 1/1 to: AL536069 from: 1 to: 889

5 SerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPhe 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 AGCCTTGACGATGCTGTGACCATGTTGCTCTCTCTGTGAGCAGCATGTT 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 eAlaPheGlnSerGluGlnValLeuAlaValProAlaGlnThrSerArg 38

```

```

53 CGCGTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCCATGAAACCTCTAGC 102
38 lInValGlnValLeuGlnAsnLeuThrThrThrTyrGlnIleValLeuTrp 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 AAGTTCAAGTCTACAGAAATCTTACTACAACTATGACATTTGTTCTCG 152
55 GluProValThrAlaAspLeuIleValLysLysLysGlnValHisPhePhe 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 CAGCCGGTACAGCTGACCTTATGTGAAAGAAAAAACAAGTCCATTTT 202
71 eValAsnAlaSerAspValAlaAspAsnValLysAlaHisLeuAsnValSer 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 TGTAAATGATCTGATGTCGACAAATGTGAAGCCCATTTAAATGTGAGCG 252
88 lYleProCysSerValLeuLeuAlaAspValGluAspLeuIleGln 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 GAATTCATGCACTGCTCTTCTGGCGACGAGCTGAAGATCTTATTACAG 302
105 GlnIleSerAsnAspThrValSerProAlaSerAlaSerTyrTyrGln 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 CAGATTTCCAAACGACAGTCAGCCCGGAGCTCCGATCGTACTATGA 352
121 uGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThr 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 ACAGTATCACTCACTAAATGAATCTATTCTTGATAGATTATTAACG 402
138 lUaArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 AGAGCATCTCTATATGCTTACAAAAATCCACATTGATCTCATTTGAG 452
155 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLys 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
453 AAGTACCCACCTGATCTTTTAAAGTTCTGTGAAAAAACAAGCAGCCCA 502
171 sAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSer 188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
503 AAATGCCATATGATGATGACTGTGGAATCCATGCCAGAAATGATCTCTC 552
188 rAlaPheCysLeuTrpPheIleGlyHis..... 197
553 CTGCTTTCTGCTTGTGTGTTCAATGACCATATACATCTTATGGAGATA 602
197 ..... 197
603 ATAGGCAATATACCAATCTCTGAGCGTTTGATTTCTATGTAATGCC 652
198 .....AsnArgMetT 201
653 GGTGCTTATGTGATGTTATGACTACTCATGAAAAAAGAAATCGAATGT 702
201 rArgLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 217
703 GGAGAAAGACCGTTCTTCTATGCGAACATCATCTGCTTGTGAACAGAC 752
218 LeuAsnSerAsnPheValSerLysHisTrpCysGluGlnGlyAlaSerSer 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
753 CTGAATAGGAATTTGCTTCCAAACACTGTGTGAGGAAGTCATCCAG 802
234 rSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProG 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
803 TTCCATCATGCTCGGAACCTACTGTGACTTTATCTCTGATCAGAAACAG 852
251 lValLysAlaValAlaSerPheLeuArg 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
853 AAGTAAAGCACTGGCTAGTTCTTGAGAGA 884

seq_name: gb_est2:BG618629

seq_documentation_block:
LOCUS BG618629 788 bp mRNA linear EST 18-APR-2001
DEFINITION 602646064F1 NIH_MGC-76 Homo sapiens cDNA clone IMAGE:4767974 5',
mRNA sequence.

```

```

ACCESSION   BG618629
VERSION     BG618629.1
KEYWORDS    GI:13670000
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 788)
AUTHORS     NIH-MGC http://mgc.ncl.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM1630 row: 0 column: 15
            High quality sequence stop: 760.
            Location/Qualifiers
                1..788
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4767974"
                /clone_1lb="NIH_MGC_76"
                /lab_host="DH10B (T1 phage-resistant)"
                /note="Organ: Liver; Vector: pMDR-LIB (Clontech); Site: 1;
                SfiI (ggcgccctggcc): Site 2; SfiI (ggccatattggc); 5' and
                3' adaptors were used in cloning as follows: 5' adaptor
                sequence: 5'-CACGGCATATGAGCC-3' and 3' adaptor sequence:
                5'-ATTCTAGAGCGGAGCGGCGGCGCATATG-3' (where B = A,
                C, or G and N = A, C, G, or T). Average insert size 1.85
                kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
                by PCR. This library was enriched for full-length clones
                and was constructed by Clontech Laboratories (Palo Alto,
                CA). Note: this is a NIH-MGC Library."
BASE COUNT  222 a 157 c 169 g 240 t
ORIGIN
alignment_scores:
    Quality: 1034.00      Length: 241
    Ratio: 5.094          Gaps: 1
    Percent Similarity: 84.232      Percent Identity: 83.402
alignment_block:
US-09-980-881-2 x BG618629 ..
Align seg 1/1 to: BG618629 from: 1 to: 788
1 MettlyslencysserleualaValleuValproIleValleuphecysgl 17
|||||
66 ATAAACCTTGGACGCTTGACATGTCATGTCCTTCCTGCTGA 115
|||||
17 uGlnHlsvAlpheAlaphheglInserGlyGlnValleuAlaAlaLeuproA 34
|||||
116 GCAGCAGTGTCTGGCTTCAGAGTGGCCAAAGTTTACGCTTCCTCTCA 165
|||||
34 rGtHsErArGlnValIglValleuGlnAsnLeuThrThrThrTyrglu 50
|||||
166 GAACCTCTAGGCAAGTTCAAGTCTACAGATCTTCTACACATATGAG 215
|||||
51 lleValleuTrpGlnProValThrAlaAspLeuIleVallyslslysgl 67
|||||
216 ATTGTTCTCTGGACGCGGTACAGCTGACCTATGTGAAGAAAACA 265
|||||
67 nValHlshpHeuValAsnAlaSerAspValAspAsnValysAlaHlsl 84
|||||
266 AGTCATTTTGTGTAATGACATCTGATGTGACAAATGTGAAGCCCAT 315

```

```

84 eUaAnValserGlyIleProCysSerValleuLeuAlaAspValGluasp 100
|||||
316 TAAATGTACGCGGAATTCATGCAAGTGTCTTCTGGAGATGTGAAGAT 365
|||||
101 leuIleGlnGlnGlnIleSerAsnAspThrValserProAlaAlaSerAl 117
|||||
366 CTATATCAACAGCAGATTTCCACAGACAGATGAGCCCGGAGCTCGC 415
|||||
117 aSeTyTrpGlnGlnTrpHisSerLeuAsnGlnIleTyserTrpIleG 134
|||||
416 ATCTGACTAGACAGATGATCTGCTCAATGAATATCTATCTTGATAG 465
|||||
134 lUpheIleThrGluArgHisProAspMetLeuThrIleHisIleGly 150
|||||
466 AATTATTAAGTACAGAGCATCTGATATGCTTACAAAAATCCATTTGCA 515
|||||
151 SerSerPheGluIlyTyProLeuTyValleuIlyValserGlyIlysgl 167
|||||
516 TCCTCATTTGAGAGTACCACCTGATGTTTAAAGSTTCTGGAAAGA 565
|||||
167 uGlnThrAlaIlyAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
|||||
566 ACAAGCAGCCAAATATGCATATGATGATGACTGGAATCCATGCCAG 615
|||||
184 lUtrPileSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
|||||
616 AATGATCTCTCGCTTCTGCTGTGTGTTACAGCCATATACTCA 665
|||||
197 ..... 197
666 TTCTATGAGGATATAGGCAATATACCAATCTCTGAGGCTTGTGANT 715
|||||
197 ..... 197
716 TCATATGTTATCCAGTGTAAATGTTGAGTGTATGACTACTATGGA 765
|||||
198 .....AsnArgMetTrpArglys 203
|||||
766 AAAGCAATCGATGTGTGAGAAAG 788
seq_name: gb_est2:BI144644
seq_documentation_block:
LOCUS      BI144644                      853 bp      mRNA      linear      EST 05-JUL-2001
DEFINITION 602909332P1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5051358 5',
            mRNA sequence.
ACCESSION  BI144644
VERSION    BI144644.1 GI:14604645
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 853)
AUTHORS     NIH-MGC http://mgc.ncl.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM1139 row: 0 column: 07
            High quality sequence stop: 849.
            Location/Qualifiers
                1..853
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
FEATURES
source

```



```

/clone="IMAGE:5051358"
/lab_host="DH10B (TI phage-resistant)"
/Note="Organ: Liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NC1_CGAP Library."
BASE COUNT      243 a      199 c      194 g      217 t
ORIGIN

```

```

alignment_scores:
  Quality: 1015.50      Length: 287
  Ratio: 4.474          Gaps: 5
  Percent Similarity: 79.094      Percent Identity: 70.383

```

```

alignment_block:
US-09-980-881-2 x B1144644 ..

```

```

Align seg 1/1 to: B1144644 from: 1 to: 853

```

```

3 LeucySerLeuAlaValLeuValProileValLeuPheCysGluGlnI 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 CTTCATGGCCCTTGGAAATCCTGTAGACCATCATCCGCTAT...GAGCAGCA 48
19 sValPheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrs 36
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
49 TGCTTCGCCCTTCAGAGTGGCCAGGCTTTATCTGCTTCACGAAGACCT 98
36 errrpglnvalglnvalleuglnasleuthrthrtrptryglnuileval 52
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
99 CCAAGGCAAGTCACTACTAGATCTTACTACTACAGCTATGAGCTGCT 148
53 LeuTrpGlnProValThrAlaAspLeuIleValIleLysLysGlnValHI 69
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
149 CTCTGGAGCAGCAGCAGCAGCAGTGAATCATCGAAGAAAGAAAGACCA 198
69 sphephevalasnaIaseraspvalaspaasvalIysAlaHisLeuasn 86
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
199 CTTTTTGTAAATGCGTCTGATGCGACAGTGCACAAAGCCCATTTAATG 248
86 alserglyIleProCysSerValleuLeuAlaAspValGluAspLeuIle 102
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
249 TGACGAGAAATTCATTACGTTCTGATGAAACAGCGTGGAGACTTAAT 298
103 GlnGlnGlnIleSerAsnaSpThrValSerProArgAlaSerAlaSerTy 119
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
299 GAACAGCAGACTTTCATGACAGCGTCAAGCCCGCGCTCCGCTCATTA 348
119 rTyrglnGlnTrpHisSerLeuasnGluIleTyrsertTrpIleGluPhe 136
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
349 CTATGACAGATATCATCTCGCTAAATCTATCTCTGATAGAGTCA 398
136 lethrGlnUAghIspProasmetleuthrLysIleHisIleGlySerSer 152
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
399 TAACTGACACGACATCTGACATGCTCCAGAAATCTACATCGATCATCA 448
153 PheGlnLysTyrrProLeuTyrrValleuLysValSerGlyLysGlnGln 169
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
449 TTGAGAGAAATGACCATTTATGTTTAAAGGTCTCAGAAAGAAAGAAAG 498
169 rAlaLysAsnaAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrp 186
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
499 AATCAAAATGACATCTGATGATCGACTGTGAATCCATGCCAGAAATGGA 548
186 leserProAlaPheCysLeuTrpPheIleGlyHis..... 197
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
549 TTTCACCTGCTTTCTGTGTTGTTAGGCTTACGCTACGACAAATTCAT 598
197 ..... 197
599 GGGAAAGAAATCTGTATACAGACTTCTGAGGACGATGATTTCTACAT 648

```

```

198 .....Asna 199
649 CATGCCCGTATGAACGTGATGGCTATGACTACACGTGAAAAAGATTC 698
199 rgmetTrpArgLysAsnaArgSerPheTyrrAlaAsnaHisCysIleGly 215
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
699 GAATCTGAGAGAAACCGT...CTGTCAACAAGAACCAACCTGGCTGGG 745
216 ThrAspLeuAsnSerAsnPhenValSerLysHisTrpCysGluGlnGlyAl 232
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
746 CCAAGACCTGACAGAGAACTTCTCCAAACACAGTGTGAGAAAGGTGC 795
232 aserSerSerSerCysSerGluThyTrpCysGlyLeuTrpProGluSe 248
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
796 GTCCAGTTCCTCCTGCTGTGAACCTACTGTGGGACTTATCTGTGATC 845
248 rGluPro 250
| | | | |
846 TGAGCCG 852
seq_name: gb_est2:BG616456

```

```

seq_documentation_block:
LOCUS      BG616456              738 bp      mRNA      linear      EST 18-APR-2001
DEFINITION 602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',
ACCESSION   BG616456
VERSION     BG616456.1 GI:13667827
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 738)
REFERENCE   NIH-MGC http://img.nci.nih.gov/
AUTHORS    National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
COMMENT     CDNA library preparation: CLONETECH Laboratories, Inc.
            Tissue Procurement: CLONETECH Laboratories, Inc.
            CDNA library Arrayed by: rpe I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCMI595 row: c column: 05
            High quality sequence stop: 666.
            Location/Qualifiers

```

```

FEATURES
    source
        1..738
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4733500"
            /lab_host="DH10B (TI phage-resistant)"
            /Note="Organ: liver; Vector: pDMR-LTB (Clontech); Site_1:
            SfiI (ggcgctcgcc); Site_2: SfiI (ggccatagcc); 5' and
            3' adaptors were used in cloning as follows: 5' adaptor
            sequence: 5'-CACGGCATTTATGGCC-3' and 3' adaptor sequence:
            5'-ATTCTAGAGCCGACGGCCGACATG-dt(30)BN-3' (where B = A,
            C, or G and N = A, C, G, or T). Average insert size 1.85
            kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
            by PCR. This library was enriched for full-length clones
            and was constructed by Clontech Laboratories (Palo Alto,
            CA). Note: this is a NIH_MGC Library."

```

```

BASE COUNT      209 a      161 c      153 g      215 t
ORIGIN
alignment_scores:
  Quality: 996.50      Length: 296
  Ratio: 4.449          Gaps: 4
  Percent Similarity: 75.676      Percent Identity: 69.257

```


184 LUTRPILESERPROLAPHESYLEUTRIPHELLEGLYHIS 197
 ||||||||||||||||||||||||||||||||||||||||
 613 AATGATCTCTCTCGCTTCTGTTGTTGATGACATGACCAT 653

seq_name: gb_est2:BG569281

seq_documentation_block:

LOCUS BG569281 852 bp mRNA linear EST 10-APR-2001
 DEFINITION 602988573F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722354 5',
 mRNA sequence.
 ACCESSION BG569281
 VERSION BG569281.1 GI:13576934
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 852)

NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LICM1580 row: b column: 19
 High quality sequence stop: 751.

FEATURES

source

Location/Qualifiers
 1..852

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4722354"
 /clone_lib="NIH_MGC_76"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggccgctcgagc); Site_2: SfiI (ggccatagcgc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATATAGGCC-3' and 3' adaptor sequence:
 5'-ATCTAGAGCGCCGAGCGCCGACATG-drf(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH-MGC Library."

BASE COUNT 251 a 177 c 187 g 237 t

ORIGIN

alignment_scores:

Quality: 965.00 Length: 260
 Ratio: 4.825 Gaps: 5
 Percent Similarity: 76.923 Percent Identity: 74.231

alignment_block:

US-09-980-881-2 x BG569281 ..

Align seg 1/1 to: BG569281 from: 1 to: 852

157 ProleuTYValleuLYValSerGlySGluGlnThrAlaIysAsnAl 173
 ||||||||||||||||||||||||||||||||||||||||
 2 CCACTCTATGTTTAAAGTTCTGGAAAACAGCAAGCAAAAATGC 51
 ||||||||||||||||||||||||||||||||||||||||
 173 AileTrPIleAspCysGlyIleHISAlaArgLUTRPIleSerProAlaP 190
 ||||||||||||||||||||||||||||||||||||||||
 52 CATATGATTCACGTGGAATCCATGCCAGCAATGATCTCTCTGCTT 101
 ||||||||||||||||||||||||||||||||||||||||
 190 hecysLeuTrPheIleGlyHIS..... 197
 ||||||||||||||||||||||||||||||||||||||||

102 TCTGCTTGTTGATCATAGGCCATATACTCATTTGATGGATTAATAGG 151
 ||||||||||||||||||||||||||||||||||||||||
 197 197
 152 CAATATACCAATCTCTGAGCTTGATTTCTATGTTATGCCAGTGGT 201
 ||||||||||||||||||||||||||||||||||||||||
 198 AsnArgMetTrp.Arg 202
 202 TAATGTGATGATGTTATGATGACATCAGGAAAAAGCAATCGAATGGAGCA 251
 ||||||||||||||||||||||||||||||||||||||||
 203 LysAsnArgSerPheTYrAlaAsnAsnHIScysIleGlyThrAspLeuAs 219
 ||||||||||||||||||||||||||||||||||||||||
 252 AAGAACGGTTCTTTCTATGCGAACAATCATTTGATCGGAACAGACCTGAA 301
 ||||||||||||||||||||||||||||||||||||||||
 219 nSerAsnPheValSerLysHISTrpCysGluGluGlyAlaSerSerSers 236
 ||||||||||||||||||||||||||||||||||||||||
 302 TAGGAACCTTTCCTCCAAACACTGTGTGAGGAAGGTGATCCAGTTCTCT 351
 ||||||||||||||||||||||||||||||||||||||||
 236 erCysSerGluThrTYrCysGlyLeuTYrProGluSerGluProGluVal 252
 ||||||||||||||||||||||||||||||||||||||||
 352 CATGCTCGGAAACCTACTGTGGACTTTATCCGTAGACAGCAACAGACATG 401
 ||||||||||||||||||||||||||||||||||||||||
 253 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTy 269
 ||||||||||||||||||||||||||||||||||||||||
 402 AAGCGAGTGTAGTATTCTTGTAGAGAAATATCAACCAAGATTAAGCATTA 451
 ||||||||||||||||||||||||||||||||||||||||
 269 rIleSerMetHISerTYrSerGlnHISLeuAlaPheProTYrSerTYrT 286
 ||||||||||||||||||||||||||||||||||||||||
 452 CATCACATGCAATTCATCTCCAGCATATAGTGTTCATATTCCTATA 501
 ||||||||||||||||||||||||||||||||||||||||
 286 hArgSerLysSerLysAspHISGluGluLeuSerLeuValAlaSerGlu 302
 ||||||||||||||||||||||||||||||||||||||||
 502 CACGAAGTAAAGCAAGACATGAGAACTGCTCTGATACCAGTGA 551
 ||||||||||||||||||||||||||||||||||||||||
 303 AlaValArgAlaIleAspLysThrSerLysAsnThrArgTYrThiSGL 319
 ||||||||||||||||||||||||||||||||||||||||
 552 GAGGTTCGGTATTGAGAAATATTAGTAAATATACAGTATACACATGCG 601
 ||||||||||||||||||||||||||||||||||||||||
 319 YHISGlySerGluThrLeuTYrLeuAlaProGlyGlyLysAspAspTYPI 336
 ||||||||||||||||||||||||||||||||||||||||
 602 GCATGGCTCAGAAACCTTATACCTAGCTTCCTGAGGTGGGAGCAATGGA 651
 ||||||||||||||||||||||||||||||||||||||||
 336 LeTYr.AspLeuGlyIleLysTYrSerPheThr..... 346
 ||||||||||||||||||||||||||||||||||||||||
 652 TCTATGACATTTGGCATAGATATTCGTTACAATGGAACCTTCAGATA 701
 ||||||||||||||||||||||||||||||||||||||||
 347SerAsn..ProProVa 351
 ||||||||||||||||||||||||||||||||||||||||
 702 CGGGCACAATACGGAATTTCTTGCTGCCGAGCGTTACATCAACACACCTGT 751
 ||||||||||||||||||||||||||||||||||||||||
 351 IGLuLysLeuLeuProLeuSerLeu 359
 ||||||||||||||||||||||||||||||||||||||||
 752 AGAGAACTTTTGGCCGCTGTCTCTTA 776
 ||||||||||||||||||||||||||||||||||||||||

seq_name: gb_est2:BI332659

seq_documentation_block:

LOCUS BI332659 1022 bp mRNA linear EST 30-JUL-2001
 DEFINITION 602984408F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5137369 5',
 mRNA sequence.

ACCESSION BI332659

VERSION BI332659.1 GI:15017316

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 1022)

NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed By: The I.M.A.G.E. Consortium (LMLT)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMLT at:
<http://image.llnl.gov>

Plate: L1M11337 row: 0 column: 02
High quality sequence stop: 841.

FEATURES
Location/Qualifiers

1..1022
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5137369"
/clone_lib="NCI_CGAP_L19"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 303 a 233 c 225 g 261 t
ORIGIN

alignment_scores:

Quality: 957.50 Length: 336
Ratio: 3.845 Gaps: 7
Percent Similarity: 74.107 Percent Identity: 60.714

alignment_block:

US-09-980-881-2 x B1332659 ..

Align seg 1/1 to: B1332659 from: 1 to: 1022

```

3 Leucylserinealavalvalleuvalprolleuvalphecysgluinh1 19
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
3 CTTGATGGCTTGGAATCTGTGTAGCCATCATCTCTAT...GAGCAGCA 49
19 svalphealaphelinserserlyglnvalleualalaleuProargThrs 36
| | ::::::::::::::::::::| ::::::::::::::| ::::::::::|
50 TGGCTTGCTTCAGAGTGGCCAGGTTTATCTGCTCTCCAGAAACCT 99
36 erarglnvalglnvalleuGlnAsnLeuthrThrTyrgluileval 52
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
100 CCAAGCAAGTTCACACTTCTGACATCTTACAAACGTATGAGGTGCTT 149
53 LeutrglnProvalThrAlaAspleuilevallysylsnglnvalh1 69
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
150 CTCTGGCAGCCAGTACAGCTGAATTCATGAGAGAAAGAAAGATCCA 199
69 spherhevalasnalaseraspyalaspasvallysalahisLeuAsnV 86
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
200 CTTTTCGTGAATGCGTGTGATGTCAGACAGTCAAGCGCATTTAATG 249
86 alserglylieProCysSerValleuLeuAlaAspvalGluAspleuile 102
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
250 TGAGCAAAATTCATTTACGTTCTGTGTAACAACGTGAGAGACCTAAT 299
103 GlnGlnGlnIleSerAsnAspThrValSerProargAlaSerAlaSerTy 119
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
300 GAACAGCAGACTTTCATTAAGACAGCGTACGCCGCCCTCCGCTTCATA 349
119 rTyrgluGlnIleThrHisSerLeuAsnGlnIleTySerTrpIleGluPhe 136
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
350 CTATGACAGATATCACTCGTAATGAATCTATCTGTGTAATAAGTCA 399
136 lethrgluatghisProAspMetLeuthrTyrIleHisIleGlySerSer 152
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
400 TAACTGAACAGCATCTGACATGCTCCAGAAATCTACATCGATCATCA 449
153 PheGluIuTySerProLeuTyValleuLeuValSerIlysgluGlnTh 169
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|

```

```

450 TTCGAGAGATCCACCTTATGTTTAAAGCTTCAGAAAGCAACAAG 499
169 rAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgIuTrpI 186
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
500 AATCAAAATGCAATTCGATGCACTGTGGAATCATGCCAGAAATGGA 549
186 leSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
550 TTTCACTGCTTCTGTTG..TGGTTCATAGGCTACGTACACAAATTCAT 598
197 ..... 197
599 GGGAAAGAAATCTGTATACCACTCTGAGCAGCGTGAATTTCACA 648
198 .....A 198
649 TCATGCCCGTATGGAACGTGATGGCTATGATACATACAGTGGACAAAGA 698
198 snATGmetTrpArgLysAsnArgSerPheTyralaAsnAsnHisCysIle 214
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
699 ATCGAATGTTGCAGAGAACCGTCTGCTCACAGAACACACCGCTGCTGT 748
215 GlyThr...AspleuAsnSerAsnPheValSerIlyshIstPrCysGluG1 230
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
749 GGGCAACATGACTGACAGAGACTGCTTCACAAACAGTGTGAGAA 798
230 uGlyAlaSerSerSerSerCysSerGluThrTyrcysGlyLeuTy..Pro 246
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
799 AGGTTGGCTCAAGTTCTCTCTGCTGCTGAACTACTGTGATTTATTC 848
247 GluSer..GluProGluValLysAlaValAlaSerPheLeuArgAsnI 263
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
849 GAGTCTAGAGCCAGACAGTGAAGCCAGCGGTCTTTTGAGAAATAT 898
263 leAsnGlnIleLysAlaTyrlleSerMetHisSerTyrcysGlnHisIle 279
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
899 CTGACCCCTTAAGTT.....ACTTCAAGATGGCTC 930
280 ValPhePro.....TyrSerTyrrhrArgSerIlySerIlyAspH1 293
||| ::::::::::::::::::::| ::::::::::::::| ::::::::::|
931 ATAATTCACAAATATGTTCTTATTCATATATAGAGAAACAGAGAGCA 980
293 sglu 294
|||
981 AGAA 984
seq_name: gb_est1:AV646979
seq_documentation_block:
LOCUS AV646979 735 bp mRNA linear EST 15-JAN-2002
DEFINITION AV646979 G1C Homo sapiens cDNA clone G1CATD06 3', mRNA sequence.
ACCESSION AV646979
VERSION AV646979.1 GI:9867993
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 735)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z., and Han,Z.
Insight into hepatocellular carcinoma carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
JOURNAL MEDLINE
CONTACT Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)

```

Fax: 86-21-50801922
Email: hanzge@nc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers

FEATURES

1..735

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCMTD06"
/clone_id="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 200 a 158 c 172 g 205 t
ORIGIN

alignment_scores:
Quality: 949.00 Length: 193
Ratio: 5.048 Gaps: 1
Percent Similarity: 97.409 Percent Identity: 96.891

alignment_block:
US-09-980-881-2 x AV646979

Align seg 1/1 to: AV646979 from: 1 to: 735

```

6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAl 22
7 CTTGCGGCTCTTGTACCCATTGTTCTCTGTGACAGACAGATGCTCTCGC 56
22 apheGlnSerGlyGlnValLeuAlaLeuProAArgThrSerArgGln 39
57 GTTTCAGAGTGGCCAAAGTCTAGCTCTCTCTAGAACCTCTAGGCAAG 106
39 aAGlnValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTyrGln 55
107 TTCAGTCTTACAGATCTTACACAAATATGAGATTGTTCTCTGCGAG 156
56 ProValThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 72
157 CCGGTAACGGTGACCTTATGTGAAGAAAAAACAAGTCCATTCTTTTGT 206
72 IAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGly 89
207 AATGATGATGATGTCGACAAATGTGAAGCCCAATTAATGTGAGCGGAA 256
89 IeProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 105
257 TTCATGCGAGTGTCTGTGGCAGAGCTGGAAGATCTTATTCACAGCAG 306
106 ILeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 122
307 ATTTCCAAAGCACAGTCAGCCCGAGGCTCCGCATCGCATATGACAA 356
122 nTyrHisSerLeuAsnGluIleTyrSerTyrPheIleupheIleThrGlu 139
357 GATACACTACATAAATGAATCTATCTTGGATAGAAATTTAATACCTG 406
139 rGHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 155
407 GGCATCTGATGCTGTACAAAATCCACATGATGATCCATTTGAGAAG 456
156 TyrProLeuTyrValLeuLys ValSerGlyLysGluGlnThrAlaLys 172
457 TACCCACTCTATGTTTAAAGGGTGTGGAAAGAACAAACAGCCAA 506
172 snAlaIleTyrPheAspCysGlyIleHisAlaArgGluTyrPheSerPro 188
507 ATGCCATATGATGATGCTGGAATCCATGCCAGAAATGATCTCTCT 556
189 AlaPheCysLeuTyrPheIleGlyHis 197

```

557 GCTTCTGTTGGGGGTCAATAGGCCAT 583

seq_name: gb_est2:BG618813

seq_documentation_block:

LOCUS BG618813 750 bp mRNA linear EST 18-APR-2001
DEFINITION 602646186F1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4767885 5',
MRNA sequence.
ACCESSION BG618813
VERSION BG618813.1 GI:13670184
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 750)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov.
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
plate: LNCM1630 row: k column: 22
High quality sequence stop: 622.

FEATURES

source

1..750
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4767885"
/clone_id="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcgcc); Site_2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCGCATTAATGCCC-3' and 3' adaptor
sequence: 5'-ATTCTAAGGCCGAGGCGGAGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

BASE COUNT 219 a 146 c 171 g 213 t
ORIGIN

alignment_scores:
Quality: 927.50 Length: 210
Ratio: 4.781 Gaps: 5
Percent Similarity: 92.381 Percent Identity: 90.476

alignment_block:
US-09-980-881-2 x BG618813

Align seg 1/1 to: BG618813 from: 1 to: 750

```

1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysG 17
11 ATGAAGCTTTGCAGCTTCGAGTCTTGTACCCATTGTTCTCTGTGA 60
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
61 GCAGCATCTCTTCGGTTTCAGAGTGGCCAAAGTCTACAGCTCTTCC 110
34 rGThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyrG 50
111 GAACCTTAGCAGAGTTCAGTCTTACAGATCTTACAGACATATAGAG 160

```

```

51 11leValleuTrpGlnProValThrAlaAspLeuIleValylsLysGlu 67
|||||
161 ATGCTCTCTGGCGACCGGTACAGCTGACCTTATGTGTGAAGAAAAACA 210
67 nvalHisPhePheValAsnAlaSerAspValAspAsnValLysLysLys 84
|||||
211 AGTCATATTTTGTAAATGATCTGATGTCGACAAATGTGAAGCCCATTT 260
84 euAsnValSerGlyIleProCysSerValIleuAlaAspValGluAsp 100
|||||
261 TAATGTGACGCGAATCCATGCAGTGTCTGCTGGCAGAGTGGAGAT 310
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
311 CTATATCAACAGCAGATTTCCACAGACAGTCAAGCCCGAGCCCTCGC 360
117 AserTyTrpGluGlnIleThrHisSerLeuAsnGlnIleTySerTrpIleG 134
|||||
361 ATGCTACTATGACAGATATCATCTCATTAAGAAATCTATCTTGATAG 410
134 IupheIleThrGluArgHisProAspMet.LeuThrLysIleHisIleG 150
|||||
411 AATTATTAAGTGAAGGACATCCATATGCTTACAAAATCCACATTTGG 460
150 Y.SerSerPheGluLysTrpProLeu.TyrValIleuLysValSer.GlyL 166
|||||
461 CATCTCATTCGAGAGAACTAACACAGTCGATGTTTAAAGTTCTCGGAA 510
166 ysgIuGlnIleThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
|||||
511 AAGAACAGACAGCGCAAAATGCCATATGATGATGATGATGATGATGCC 560
183 ArgGluTrpIleSerProAlaPheCysLeuTrp..... 193
|||||
561 AGAATAATGATCTCTCTGCTTCTGCTGTGTGATAGCCCATATTAAT 610
194 ....PheIleGlyHisAsnArg 199
|||||
611 CAATTTCTATATGGATATAG 632
seq_name: gb_estc2:BG565100

seq_documentation_block:
LOCUS BG565100 747 bp mRNA linear EST 10-APR-2001
DEFINITION 602583720F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711348 5',
mRNA sequence.
ACCESSION BG565100
VERSION BG565100.1 GI:13572753
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 747)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: CLOMTECH Laboratories, Inc.
cDNA Library Preparation: CLOMTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM151 row: h column: 05
High quality sequence stop: 714.
Location/Qualifiers
1. 747
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source

```

```

/clone="IMAGE:4711348"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1;
SfiI (ggcgctggcc): Site 2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATAGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGAGCGCGCCAGCAGT-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

BASE COUNT 220 a 154 c 166 g 207 t
ORIGIN

alignment_scores:
Quality: 902.00 Length: 251
Ratio: 4.422 Gaps: 6
Percent Similarity: 81.275 Percent Identity: 77.291

alignment_block:
us-09-980-881-2 x BG565100 ..

Align seg 1/1 to: BG565100 from: 1 to: 747

52 ValLeuTrpGlnProValThrAlaAspLeuIleValylsLysGluVa 68
|||||
1 GTTCTCTGCGACCGCGTACAGCTGACCTTAT.GTGAAGAAAAACAAGT 49
68 1HisPhePheValAsnAlaSerAspValAspAsnValLysLysLys 85
|||||
50 CCATTTTGTGAGATGATCGATGATGATGATGATGATGATGATGATGAT 99
85 snValSerGlyIleProCysSerValIleuAlaAspValGluAsp 101
|||||
100 ATGTGAGCGGAATCCATGACAGTGTCTGCGCAGATGTGAAGATCTT 149
102 IlegInGlnGlnIleSerAsnAspThrValSerProArgAlaSerAla 118
|||||
150 ATTCACACGCGATTTCCACAGCAGATCAGCCCGAGCCCTCGCATC 199
118 rTyTrpGluGlnIleThrHisSerLeuAsnGlnIleTySerTrpIleG 135
|||||
200 GTACATGACAGATATCATCTCATTAAGAAATCTTGTGAGAGAT 249
135 heIleThrGluArgHisProAspMet.LeuThrLysIleHisIleGlySer 151
|||||
250 TTATTAAGTGAAGGACATCCATATGCTTACAAAATCCACATTTGATCC 299
152 SerPheGluLysTrpProLeuTrpValIleuLysValSerGlyLysGlu 168
|||||
300 TCATTGGAAGTACACCTCATGTTTAAAGTTTGTGGAAGAAACA 349
168 nThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGlu 185
|||||
350 AGCAGCCCAAAATGCCATATGATGATGATGATGATGATGATGATGAT 399
185 rPheIleSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
|||||
400 GGACCTCTCTGCTTCTGCTGTGTGATGATGATGATGATGATGATGAT 449
197 ..... 197
|||||
450 TATGGATTAATAGGCAATATACATCTCTGAGGCTTGTGATTTCTA 499
198 .....AsnArgMetTrpArg.... 202
|||||
500 TGTATGCCAGTGTATATGATGATGATGATGATGATGATGATGATGAT 549
203 .....LysAsnArgSerPheTyTrpAlaAsnHisGly 213
|||||

```

550 AATCGAATGTGGAGACAGACCGTTCTTCTATGGACATCATTC 599
 214 TlleglThraspleuasenseraenphelValSerlyshstPcysglucl 230
 600 ATCGGACACACCTGAAATAGCAACTT.GCTTCCAAACACGCTGTGAGGA 648
 230 u.glyalaserSerSer.CysSerGlu.ThlTy.CysGlyLeuTy 245
 649 AGGCTGATCCTACTTCTCTGCGAAACCTACTTGTGACCTTAT 698
 246 ProgluSerGluProgluValLysAlaValAlaSerPheLeuArg 260
 699 CCGAGTCAGACACGAGTG.GAAGCAGTGGCTAGTTCTTGAAG 742
 seq_name: gb_est2:BG568240

seq_documentation_block:
 LOCUS BG568240 691 bp mRNA linear EST 10-APR-2001
 DEFINITION 602587149F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4715797 5',
 mRNA sequence.
 ACCESSION BG568240
 VERSION BG568240.1 GI:13575893
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 691)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgepds-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMU at:
 http://image.llnl.gov
 plate: L1CM1563 row: a column: 14
 High quality sequence stop: 691.

FEATURES
 source
 1..691
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4715797"
 /clone_id="NIH_MGC_76"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggccgcctggcc); Site_2: SfiI (ggccatattggc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGCCGAGCGCCGACATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."

BASE COUNT 204 a 143 c 144 g 200 t
 ORIGIN

alignment_scores:
 Quality: 898.00 Length: 232
 Ratio: 4.582 Gaps: 4
 Percent Similarity: 84.483 Percent Identity: 81.034

alignment_block:
 US-09-980-881-2 x BG568240 ..
 Align seg 1/1 to: BG568240 from: 1 to: 691

1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysgl 17
 19 ATGAGCTTTGGACGCTTGGAGTCTTGTACCCATTGTTCTTCTGTGA 68
 17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaIleuProa 34
 69 GCAGCATGCTTCCGCTTTCAGATGAGTGGCCAACTTACGCTCTTCCGA 118
 34 rGthSerArgGlnValGlnValLeuGlnAsnLeuThrThrTrpArg 50
 119 GAACCTTACGCAAGTTCAGATTCTACAGAACTTCTACACATATGAG 168
 51 TleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysgl 67
 169 ATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTAT.GTGAGAGAAAAACA 217
 67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHis 84
 218 AGTCCATTTTGTGTAATGATCATGTGATGCGAATGTGMAAGCCCAT 267
 84 euAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluasp 100
 268 TAAATGTGAGCGGAATTCATGACATGCTCTTGTGCGACAGCTGGAAGT 317
 101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
 318 CTATATCAACAGCAGATTCTCAACGACACAGTACAGCCCGGACCTCCGC 367
 117 aSerTrpTrpGlnGlnTrpHisSerLeuAsnGlnIleTrpSerTrpIleG 134
 368 ATCTGATCTATGAACAGTATCATCTCACTAATGAATGAATCTTGTGATG 417
 134 lPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
 418 AATTATTAATCGAGAGCATTCGATATCTTACAAATAATCCACANTGGA 467
 151 SerSerPheGlnLysTrpProLeuTrpValLeuLysValSerGlyLysgl 167
 468 TCCTCATTTTGAAGATGACACCTCATGTGTTAAAGGTTTCGAAAAAGA 517
 167 uGlnThrAlaLysAsnAlaIle..... 174
 518 ACNAGAGCCCAAAATGCCATATAACTAATCTATGAGGATATAGGCC 567
 175TrrPheAspCysGlyIleHisAlaArgGlu 184
 568 AATATACCAATCTCCTGAGGCTTGTGATTC..... 599
 185 TrpIleSerProAlaPheCysLeu.....TrrPheI 195
 600 TATGTTATGCCGCTGTAATGTGATGGTATGACTACTCATG..... 644
 195 eGlyHisAsnArgMetTrpArgLysAsnArgSerPheTrpAlaAsn 210
 645 .AAACGAAATCGAATGTGAGAAACACCGCTTCTTCTATGCGAAG 689
 seq_name: gb_est1:AI048167

seq_documentation_block:
 LOCUS AI048167 724 bp mRNA linear EST 08-JUL-1998
 DEFINITION ud71b12.y1 Sugano mouse liver m1a Mus musculus cDNA clone
 IMAGE:1451327 5' similar to TR:015114 Q15114 PREPRO-PLASMA
 CARBOXYPEPTIDASE B. / mRNA sequence.

ACCESSION AI048167
 VERSION AI048167.1 GI:3296454
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 724)
 AUTHORS Maira,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., De,M., Martin,J., Morris,M.,

SOURCE

```
1. .583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCAFH12"
/clone_1pb="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
```

BASE COUNT 161 a 130 c 116 g 176 t
ORIGIN

alignment_scores:

Quality: 868.00 Length: 197
Ratio: 4.932 Gaps: 1
Percent Similarity: 89.340 Percent Identity: 87.817

alignment_block:

US-09-980-881-2 x AV693037 ..

Align seg 1/1 to: AV693037 from: 1 to: 583

```
1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCysG1 17
|||||
16 ATGAAGCTTTCAGCCCTTGACGCTCTGTAACCATGTTCTCTCTGTGA 65
|||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
|||||
66 GCAGCATGCTTCGCCCTTCACAGTGGCCAGTTCTAGCTGCTTCTCTA 115
|||||
34 rGhrSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrglu 50
|||||
116 GAACCTCTAGGCAGATTCAGTCTACAGATCTTACTACAAACATATGAG 165
|||||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
|||||
166 ATTGTTCTCTGGCAGCGGCTTACAGCTTATTTGTGAAGAAAAACA 215
|||||
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
|||||
216 AGTCCATTTTTCATTAATGCATCTGATGTCGCAATGTGAAGCCCAT 265
|||||
84 euAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
|||||
266 TAAATGTGAGCGAATTCATGCGATGCTGCTGCGCAGCGTGAAGAT 315
|||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
316 CTATATTCACACAGCAGATTTCCACAGCACAGTCAGCCCCGAGCCCTCCG 365
|||||
117 aSerTyTyrgLugLntyrHisSerLeuAsnGluIleTySerTrpIleG 134
|||||
366 ATCGTACTATGAACAGTATCAGTCACTAATAATGAATCTATCTTGGATAG 415
|||||
134 LuPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
|||||
416 AATTATTAAGTAGAGGAGCATCTGATATGCTTACAAAAATCCACATTTGGA 465
|||||
151 SerSerPheGluLysTyrrProLeuTyrrValLeuLysValSerGlyLysG1 167
|||||
466 TCCTCATTTTGAAGATGACCATCTATGTTTAAAGGTTTCTGGGAACCA 515
|||||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
|||||
516 GCC..... 518
184 LuTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
:::|||||
519 .....ACTTCTGCTGTGTGTTCAATAGGCAT 545
```


FT METAL 302 302 ZINC (BY SIMILARITY).
 FT ACT_SITE 376 376 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 171 184 BY SIMILARITY.
 FT DISULFID 243 266 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 SQ SEQUENCE 415 AA; 47515 MW; 8EA06CCADE30B6F2 CRC64;

Query Match 38.5%; Score 691.5; DB 1; Length 415;
 Best Local Similarity 40.2%; Pred. No. 8.7e-52;
 Matches 145; Conservative 59; Mismatches 114; Indels 43; Gaps 6;

QY 1 FOSGCVLALPRTSRQVOYLNTTTELYLMQPTADLYKKKQVHFVNASDVNDKA 60
 DB 20 FDGNVAVYSVGHEDHVLNIDELANTKEIDFMKPDASQVAPLTTVDYFVNAEDVADVEN 79
 QY 61 HLNVSIGPVSLLADVEDLIDQOISNDTVSPRASASYEOYHSLNEIYSWIEFITERHPD 120
 DB 80 FLEENEVHYEVLISVNRNALSQDFDSHT--RASGSHSTYKTKNKMETIEMWIOQVATDNDP 136
 QY 121 MLTKHIGSSFEKPYLYLVKSGKEQFANNAIWDGIGHAREWISPAFCLMFT----- 173
 DB 137 LVTSVIGTTEGGRNMYLKI-GKTRPNKPAIFIDCGFHAREWISPAFCQMFVREAVRTY 195
 QY 174 -----NMWRKRNSFYANNHCIGTDLNRNFASK 203
 DB 196 NOEIHMKQLDELDFYVLPVYNIDGIYTTWKDRMKRSTMTAGSSCLGVAPNKNF-NA 254
 QY 204 HWCERGAASSSCSEYTCGLYPESEPEYKAVASFLRRINQKAYISMYSQHIYFYSY 263
 DB 255 GWCERGAASRSCSEYTCGPAPSESEKETAADFIKNNLSTIKAVLTHISYQMLPYRSY 314
 QY 264 TRSKSHOEELSLVASEVRAIEKTSKRTYTHGSEITLVLGAGGDWLYDGIKYSF 323
 DB 315 DYKLPENTEELNALVAKAKEL-ATLHGTYKTYGPAATTIYPAGGSDWSDGIXKYSF 373
 QY 324 T 324
 DB 374 T 374

RESULT 2
 CBPB_HUMAN STANDARD; PRT; 417 AA.
 ID CBPB_HUMAN
 AC P15086; 060834;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein) (PASP).
 GN CPB1 OR CPB OR PCPB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NC NCBI_TaxID=9606;
 RX MEDLINE=92129345; PubMed=1370825;
 RA Yamamoto K.K., Poussette A., Chow P., Wilson H., el Shami S., French C.K.;
 RA "Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of pancreas-specific protein as pancreatic procarboxypeptidase B";
 RL J. Biol. Chem. 267:2575-2581(1992).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RC MEDLINE=98182241; PubMed=9524066;
 RA Aloy P., Catasus L., Villegas V., Reverter D., Vendrell J., Aviles F.X.;
 RA "Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B.";

RL Biol. Chem. 379:149-155(1998).
 RN [3]
 RP SEQUENCE OF 16-43.
 RC TISSUE=Pancreas;
 RX MEDLINE=89153096; PubMed=2920728;
 RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
 RT "Purification and properties of five different forms of human procarboxypeptidases";
 RL Eur. J. Biochem. 179:609-616(1989).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-lysine(or L-arginine) + H(2O) = peptide + L-lysine(or L-arginine).
 CC -1- TISSUE SPECIFICITY: PANCREAS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE ZINC CARBOXYPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb.ch).
 CC -----
 DR EMBL; M81057; AAA66973.1; .
 DR EMBL; AJ224866; CAA12163.1; .
 DR PIR; S02812; S02812.
 DR PIR; A42332; A42332.
 DR HSSP; P09955; INSA.
 DR MEROPS; M14.003; .
 DR MIM; 114852; .
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; zn_carboxypept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; zn_carboxypept; 1.
 DR PRINTS; PR00765; CARBOXYPTASEA.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 DR KEGG; K01133; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL 1 15
 FT PROPEP 16 110
 FT CHAIN 111 417
 FT METAL 176 176
 FT METAL 179 179
 FT METAL 304 304
 FT ACT_SITE 378 378
 FT DISULFID 173 186
 FT DISULFID 245 268
 FT DISULFID 259 273
 FT CONFLICT 16 16
 FT CONFLICT 17 17
 FT CONFLICT 37 37
 FT CONFLICT 208 208
 FT CONFLICT 245 245
 SQ SEQUENCE 417 AA; 47366 MW; BB1CF212D830305E CRC64;

Query Match 35.5%; Score 637.5; DB 1; Length 417;
 Best Local Similarity 36.0%; Pred. No. 3.7e-47;
 Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps 6;

QY 1 FOSGCVLALPRTSRQVOYLNTTTELYLMQPTADLYKKKQVHFVNASDVNDKA 60
 DB 22 FDEKVFERNVDEENHINIRELASTQIDFMKPDSTQIKHPSYDVFYKAEDEVYVEN 81
 QY 61 HLNVSIGPVSLLADVEDLIDQOISNDTVSPRASASYEOYHSLNEIYSWIEFITERHPD 120
 DB 82 VLKQNELQKVLISNRRNVENQDFDSRV--RATGSHSTYKTKNKMETIEMWIOQVATDNDP 138
 QY 121 MLTKHIGSSFEKPYLYLVKSGKEQFANNAIWDGIGHAREWISPAFCLMFT----- 173
 DB 139 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQMFVREAVRTY 197
 QY 174 -----GH-----NMWRKRNSFYANNHCIGTDLNRNFASK 203

```

Db 198 GREIOVTELLKLDYFVLPVINDGIYITWTKSRWKRTRSRHTGSSCGTDPNRNF-DA 256
QY 204 HMCCEGASSSSCSEYCYGLYPESEPEVKAVASFLRNINOIKAYISMHSYSOHIPEPYSY 263
Db 257 GWCETGASARNCDDEYCYGPAASESEKETKALADFINRNLSIKAYLTIHSGMILTPYSY 316
QY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMIYDLGIRKYSF 323
Db 317 AYKLGENNNAELNALAKATVKEI-ASLHGCTKYVYGPATYIYPAAGSDDMAYDQGIKYSF 375
QY 324 T 324
Db 376 T 376

RESULT 3
CBPB_PIG STANDARD: PRT: 401 AA.
AC P09935;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
CN CPB.
OC Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID:9623;
RN [1]
RP SEQUENCE OF 1-103.
RX MEDLINE-91208150; PubMed-2018774;
RA Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,
RA Aviles F.X.;
RT "Analysis of the activation process of porcine procarboxypeptidase B
RL Biochemistry 30:4082-4089(1991).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-38.
RX MEDLINE-85279427; PubMed-4026847;
RA Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;
RT "Sequential homologues between procarboxypeptidases A and B from
RL Biochem. Biophys. Res. Commun. 130:97-103(1985).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE-91114690; PubMed-1989878;
RA Coll M., Guasch A., Aviles F.X., Huber R.;
RT "Three-dimensional structure of porcine procarboxypeptidase B: a
RL structural basis of its inactivity.";
RN [4]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
RX MEDLINE-91027767; PubMed-2223783;
RA Vendrell J., Wider G., Aviles F.X., Wuehrich K.;
RT "Sequence-specific 1H NMR assignments and determination of the
RL pancreatic structure for the activation domain isolated from
RL Biochemistry 29:7515-7522(1990).
RN [5]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX MEDLINE-91114693; PubMed-1989879;
RA Vendrell J., Billeter M., Wider G., Aviles F.X., Wuehrich K.;
RT "The NMR structure of the activation domain isolated from porcine
RL procarboxypeptidase B.";
RN [6]
RP EMBO J. 10:11-15(1991).
RN [6]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX MEDLINE-93044373; PubMed-1422143;
RA Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,
RL Huber R., Wuehrich K.;
RT "Comparison of the NMR solution structure with the X-ray crystal
RT structure of the activation domain from procarboxypeptidase B.";

```

```

RL J. Biol. NMR 2:1-10(1992).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC Peptide + L-Lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -1- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/COB.html".
DR PIR: B29181; B29181.
DR PDB: 1PBA; 31-OCT-93.
DR PDB: 1NSA; 24-DEC-97.
DR MEROPS: M14.003; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000634; Zn_CarDopept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carDopept; 1.
DR PRINTS: PR00765; CARBOXYPTASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
DR Hydrolase: Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
KM 3D-structure.
FT PROPEP 1 95 ACTIVATION PEPTIDE.
FT CHAIN 96 401 CARBOXYPEPTIDASE B.
FT DISULFID 158 171
FT DISULFID 224 253
FT DISULFID 244 258
FT METAL 161 161 ZINC.
FT METAL 164 164 ZINC.
FT METAL 289 289 ZINC.
FT ACT_SITE 341 341
FT ACT_SITE 363 363 NUCLEOPHILE.
FT STRAND 14 17
FT HELIX 20 31
FT TURN 32 32
FT HELIX 43 45
FT STRAND 50 52
FT HELIX 61 69
FT TURN 70 72
FT STRAND 75 76
SQ SEQUENCE 401 AA; 45713 MW; 53129AF159A26348 CRC64;

```

```

Query Match 35.3%; Score 635.5; DB 1; Length 401;
Best Local Similarity 37.4%; Pred. No. 5,2e-47;
Matches 135; Conservative 64; Mismatches 119; Indels 43; Gaps 6;

QY 1 POSGOVIALAPRTRSOVYIOMITTYEIVLMQPVADLIVKKQVHFVNASDVNYKA 60
Db 7 FECEKYFRVNWEDENDISLHELASTROIDENKPDVSYTIKPRSHYDFYKADILAVED 66
QY 61 HNVSGIPCSVLADVEDLIQOQISNDTVSPRASASYEYHSLNIEYSWIEFITERHPD 120
Db 67 FLEONELQYEVLIINLRSLVLEAOFDSRC--RUTGSHSEKYNWETIEAMTEQVYSKND 123
QY 121 MTKIHGSEFEKPYLYLVKSGKEQTAKNAIWDGCHAREISAFCLMFI----- 173
Db 124 LLSRSAIGTTFGDMITLLKV-GKPSGNKPAIFMDCGFHAREISQAFQWYRAVARTY 182
QY 174 -----GH-----NRMKRSFYANNHCIGTDLNRNFAK 203
Db 183 GYEAHTEFLNDLDFVLPVINDGIYITWTKNRMRKTRSTAGSSCTGTDPNRNF-NA 241
QY 204 HMCCEGASSSSCSEYCYGLYPESEPEVKAVASFLRNINOIKAYISMHSYSOHIPEPYSY 263
Db 242 GWCETGASVNPCEYCYGSAASESEKETKALADFINRNLSIKAYLTIHSGMILTPYSY 301
QY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMIYDLGIRKYSF 323
Db 302 DYKLPENDAELNSLAKGAVKEI-ASLYGTYSYSGPSTIYIYPAAGSDDMAYNOGIRKYSF 360
QY 324 T 324
Db 361 T 361

```

```

RESULT 4
CBPC_MOUSE STANDARD: PRT: 417 AA.
ID CBPC_MOUSE
AC P15089;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90062123; PubMed=2584208;
RA Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA Serafini W.E.;
RT "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT novel member of the carboxypeptidase gene family."
RL J. Biol. Chem. 264:20094-20099(1989).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J05118; AAA37369.1; -
DR PIR: A34487; A34487.
DR HSSP: P09955; 1NSA.
DR MEROPS: M14.010; -.
DR MGD: MGI:88479; Cp33.
DR InterPro: IPR003146; Proprep_M14.
DR Pfam: PF02244; Proprep_M14; 1.
DR Pfam: PF00246; Zn_carboxpept. 1.
DR PRINTS: PR00765; CARBOXYPTASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE.
FT CHAIN 110 417 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 245 268 BY SIMILARITY.
SO SEQUENCE 417 AA; 48790 MM; A2B300A068D1BA6D CRC64;

```

Query Match 35.2%; Score 632; DB 1; Length 417;
 Best local Similarity 38.2%; Pred. No. 1, le-46;
 Matches 138; Conservative 57; Mismatches 122; Indels 44; Gaps 7;

```

OY 1 FQSGOYVLAALPRTSRQYVQLNTTYYEIVLWQPVYADLIVKKRQYHFVNASDVNVKA 60
DB 21 FDEKRYFRVQLQNEKHASYLKNLTQSIEIDFYRDAIHDAVMTYDVERVSEKESQTIQS 80
OY HUNVSGIPCSVLLADVEDLQOQIS-NDVYSPRASAAYEQAHSLSWEIYSWIFTERHP 119
DB 81 TLEQKHITHEILIHDLQELTEKQFDVKEDELAGRHS--YAKYNDMDKIYSWTEKMLEKHP 137

```

```

OY 120 DMLTKIHGSSPEKYPVLVKGSGEQTAKNAINIDGIIHAREMISPAFCIMPT----- 173
DB 138 EMVSRKIKGISTVEDNPLYVTKI-GKDGERRKAITMDGSHAREMISPAFCQMFYQATKS 196
OY 174 -GHN-----RMRKNRNSFPANNHCIGTDLRNRPAS 202
DB 197 YGKNKIMTKLDRMNFYVLPVFNVNGYIWSWTQDRMRKNRNSRQNSTCIGTDLRNRF-D 255
OY 203 KHWEEGASSSCSEETYGGLXPSEPEPKAVASFLRRINQIKAYISMHSYQHIVPYS 262
DB 256 VSMSSPNTNPKPCUNYKGPAPSESEKFKAVTNPIRSHLSIKAYITFHYSQWLLIPYC 315
OY 263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTYRTHGSEETLYALPAGGDDMYDLDIGITS 322
DB 316 YTFKLPRNHQDLKAVARIATDAL-STREYRRTYIGPIASTIYIKTSGSSLWVDLGIKHT 374
OY 323 F 323
DB 375 F 375

```

```

RESULT 5
CBPC_CANFA STANDARD: PRT: 416 AA.
ID CBPC_CANFA
AC P55261;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE membrane associated protein) (ZAP47).
GN CPB1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Flisiapedae; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Fukuko S.-I.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-Lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D78348; BAA11366.1; -
DR HSSP: P09955; 1PBA.
DR MEROPS: M14.003; -.
DR InterPro: IPR003146; Proprep_M14.
DR Pfam: PF02244; Proprep_M14; 1.
DR Pfam: PF00246; Zn_carboxpept. 1.
DR PRINTS: PR00765; CARBOXYPTASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 110 416 CARBOXYPEPTIDASE B.
FT METAL 175 175 ZINC (BY SIMILARITY).
FT METAL 178 178 ZINC (BY SIMILARITY).
FT METAL 303 303 ZINC (BY SIMILARITY).
FT ACT_SITE 377 377 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 244 267 BY SIMILARITY.

```

FT DISUFLD 258 272 BY SIMILARITY
SQ SEQUENCE 416 AA; 47595 MW; DFEI992CD52F8AB4 CRC64;

Query Match 35.1%; Score 630.5; DB 1; Length 416;
Best Local Similarity 36.6%; Pred. No. 1,5e-46;
Matches 132; Conservative 64; Mismatches 122; Indels 43; Gaps 6;

OY 1 FOSGOVLAALPRTSRQVOYLQNTTTEYLWQPVADLIYKKQVHFVNASDVNDVKA 60
DB 21 FEGEKEFRVNEDENNINLHTLASTQTQIDFWKPDSTQIKPHSTADFRKADILTVED 80
OY 61 HUNVSGIPCSVLADVEDLIQOQISNDYSPRASASYQYHSLNITYSWIEFIERHPD 120
DB 81 FIKQNELHVEVLINLRLTEGOFGRQV--PATGHSYKRYNMWETIEAMTQOVTSENDP 137
OY 121 MLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDGCIHAREWISPAFCIMFI----- 173
DB 138 LISRSISGTFEGRTIYLKLV-CKAGONKPAIFMDCGFHAREWISPAFQWIFREXAIRTY 196
OY 174 -----GH-----NRMKRKRSEFYANNHCIGTDLNRNFASK 203
DB 197 GOEIHTELLDKLDFYVLPVGNIDGYVTWTNRMRKTRSTQVGTNCVGTDEPTRNF-DA 255
OY 204 HMCCEGASSSCSEFYCGIYPESEPEVKAVASFLRNINQIKAYISMHSYQIIVPRYS 263
DB 256 GWCKIGASNNPCDEYCGPAASESEKTKALANFIRSNLSIKAYLIHSSQMLPYPSY 315
OY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSEFTLYLAPGGDDMIYDLGIRKSF 323
DB 316 DYKLENNMELNALAKATYKEL-ATLHGKITYTGPGATTIYPAAGSDMDADOGIKRYSF 374
OY 324 T 324
DB 375 T 375

RESULT 6
CBPC_HUMAN STANDARD; PRT; 417 AA.
AC P15088;
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
GN CPB3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=lung;
RX MEDLINE=90083291; Pubmed=2594780;
RA Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,
RA Austen K.F., Serafin W.E.;
RT "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT comparison of the protein with mouse mast cell carboxypeptidase A and
RT rat pancreatic carboxypeptidases";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mast cells;
RX MEDLINE=92105393; Pubmed=1729276;
RA Reynolds D.S., Gurley D.S., Austen K.F.;
RT "Cloning and characterization of the novel gene for mast cell
RT carboxypeptidase A";
RL J. Clin. Invest. 89:273-282(1992).
RN [3]
RP SEQUENCE OF 110-417 FROM N.A.
RX MEDLINE=9233165; Pubmed=1629626;
RA Natsunaki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsunaki M.,

RA Wintrob B.U., Rutter W.J., Goldstein S.M.;
RT "Human skin mast cell carboxypeptidase: functional characterization,
RT cDNA cloning, and genealogy";
RL J. Invest. Dermatol. 99:138-145(1992).
RN [4]
RP SEQUENCE OF 110-137.
RX MEDLINE=89214692; Pubmed=2708524;
RA Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintrob B.U.;
RT "Human mast cell carboxypeptidase. Purification and
RT characterization";
RL J. Clin. Invest. 83:1630-1636(1989).
CC -! CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O -> peptide + L-
CC amino acid.
CC -! SUBCELLULAR LOCATION: Secretory granules.
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M27717; AAA5652.1; -
DR EMBL; M73720; AAA59568.1; -
DR EMBL; M73716; AAA59568.1; JOINED.
DR EMBL; M73717; AAA59568.1; JOINED.
DR EMBL; M73718; AAA59568.1; JOINED.
DR EMBL; M73719; AAA59568.1; JOINED.
DR EMBL; S40234; AAB2578.2; ALT_SEQ.
DR PIR; A43929; A43929.
DR HSSP; P09955; INSA.
DR MEROPS; M14.010; -.
DR MIM; 114851; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept. 1.
DR PRINTS; PR00765; CRBOXYPRASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1 15
FT PROPEP 16 109
FT CHAIN 110 417
FT METAL 176 176
FT METAL 179 179
FT METAL 304 304
FT ACT_SITE 378 378
FT DISUFLD 173 186
FT DISUFLD 245 268
FT SEQUENCE 417 AA; 48700 MW; 8CB90DB758117B24 CRC64;

Query Match 34.6%; Score 622; DB 1; Length 417;
Best Local Similarity 35.5%; Pred. No. 7.8e-46;
Matches 128; Conservative 71; Mismatches 118; Indels 44; Gaps 7;

OY 1 FOSGOVLAALPRTSRQVOYLQNTTTEYLWQPVADLIYKKQVHFVNASDVNDVKA 60
DB 21 FEGEKEFRVNEDENNINLHTLASTQTQIDFWKPDSTQIKPHSTADFRKADILTVED 80
OY 61 HUNVSGIPCSVLADVEDLIQOQISNDYSPRASASYQYHSLNITYSWIEFIERHPD 119
DB 81 ALDQNMHVEILLHDQEEIEKQFDVKEIDPGRHS--YAKYNNWETIVAMTERMDKYP 137
OY 120 DMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDGCIHAREWISPAFCIMFI----- 173
DB 138 EWSRIKIGSTVDNPLVYAKI-GEKNERKKAIFMDCGFHAREWISPAFQWIFVQATKT 196
OY 174 -----GH-----NRMKRKRSEFYANNHCIGTDLNRNFASK 202

```

DB 197 YGNKTKMTLLDLMNFYILPEVNDGYIMSWTKNRMKRNKRNKSNKSKICGTDLNRF-N 255
OY 203 KHWCEGASSSSEFTYCGLYPESEDEPKAVASFLRNINOKAVYISMYSQHIYFPPS 262
DB 256 ASMNSIPNTNDCADNDYKRSAPSESEKETKAVTNFINSHLNEIKVYITFYSOMLLEPPG 315
OY 263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEFLYLAPEGGDWITDGIKYS 322
DB 316 YTSKLEPNHEDLAKYAKIGTDVL-STRYETRYTYGPIESTIYPIGSSSLDMAYDGIKHT 374
OY 323 F 323
DB 375 F 375

```

```

RESULT 7
CBPB_BOVIN STANDARD: PRT: 306 AA.

```

```

AC P00732: 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B (EC 3.4.17.2).
GN Cpb.
OS Bos taurus (Bovine)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP MEDLINE=75217824; PubMed=1057162;
RA Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT "Amino-acid sequence of bovine carboxypeptidase B.";
RT Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
RN [2]
RP SEQUENCE OF 31-93: 131-181: 263-265 AND 292-306.
RA MEDLINE=74260705; PubMed=483744;
RN [3]
RP "Primary structure of bovine carboxypeptidase B. Inferences from the
RT locations of the half-cysteines and identification of the active site
RT arginine.";
RT J. Biol. Chem. 249:3756-3764(1974).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RN [5]
RP MEDLINE=76265065; PubMed=957425;
RA Schmid M.F., Herriott J.R.;
RT "Structure of carboxypeptidase B at 2.8-A resolution.";
RT J. Mol. Biol. 103:175-190(1976).
RN [6]
RP ACTIVE SITE.
RP MEDLINE=70007159; PubMed=5344132;
RA Plummer T.H. Jr.;
RT "Isolation and sequence of peptides at the active center of bovine
RT carboxypeptidase B.";
RT J. Biol. Chem. 247:7864-7869(1972).
RN [7]
RP CATALYTIC ACTIVITY: Peptidyl-L-lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC PIR: A00912; CPOB.
CC DR PDB: 1CPB; 30-SEP-83.
CC DR MEROPS: M14.003; -.
CC DR InterPro: IPR000834; Zn_carpopept.
CC DR Pfam: PF00246; Zn_carpopept. 1.

```

```

DR PRINTS: PR00765; CROXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3d-structure.
FT DISULFID 63 76
FT DISULFID 135 158
FT DISULFID 149 163
FT METAL 66 66
FT METAL 69 69
FT METAL 194 194
FT ACT_SITE 246 246
FT ACT_SITE 268 268
SQ SEQUENCE 306 AA; 34612 MM; C329D2655C44A172 CRC64;
NUCLEOPHILE.

```

```

Query Match 32.6%; Score 586; DB 1; length 306;
Best Local Similarity 43.9%; Pred. No. 6,2e-43;
Matches 116; Conservative 44; Mismatches 64; Indels 40; Gaps 5;

```

```

OY 98 YEQYSLNEIYSWIEFIERHRPDMITKHISSEFKYPLVLYKSGKEQTAKNAIWDG 157
DB 6 YEKYNNMETIEMWEQVASENDPLISRSAICTTFLGNTIYLLKV-GKPGSKPAVPMDCG 64
OY 158 IHAREWISPAECLWFI-----GH-----NFMWR 180
DB 65 FHAREWISPAFCOMFVREAVRTYGREIHMTFELDKDFYVLPVYNIDGIYTTWTNNMR 124
OY 181 KNRSFYANNHICGIDLNRFNFSKHWCEGASSSCSEYTCGLYPESEPKAVASFLRN 240
DB 125 KTRSTRGSSCTGTDLNRF-DAGWCSTIGASNNPCSEYTCGSAESESERKSAVADFI RNH 183
OY 241 INQIKAVISMYSQHIYFPTSTRSKDHEELSLVASEAVRAIEKTSKNTRYTHGHS 300
DB 184 LSLIKAVYLFHSYQMMLYPSYDYKLPKNVLENTLAKAVKVL-ASLHGTYISGPGA 242
OY 301 ETLILAPEGGDWITDGIKYSFT 324
DB 243 TTYIPASGGSDMAYDGIKYSFT 266

```

```

RESULT 8
CBPC_RAT STANDARD: PRT: 309 AA.
AC P21961: 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE A3).
GN CPB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RP MEDLINE=91105153; PubMed=1988052;
RA Cole K.R., Kumar S., Le Trong H., Woodbury R.C., Walsh K.A.,
RA Neurath H.;
RT "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT enzyme activity within mast cell granules.";
RT Biochemistry 30:648-655(1991).
RN [2]
RP CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC PIR: A33118; A33118.
CC DR PIR: A38395; A38395.
CC DR HSSP: P09955; INSA.
CC DR MEROPS: M14.010; -.
CC DR InterPro: IPR000834; Zn_carpopept.
CC DR Pfam: PF00246; Zn_carpopept. 1.

```



```

DR PRINTS: PR00765: CROXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT METAL 68 68
FT METAL 71 71
FT METAL 196 196
FT ACT_SITE 248 248
FT ACT_SITE 270 270
FT DISULFID 137 160
FT VARIANT 1 1
SEQUENCE 309 AA: 35786 MW: 20330FABC3EE83EF CRC64;

Query Match 30.7%; Score 552; DB 1; Length 309;
Best Local Similarity 42.6%; Pred. No. 5.1e-40;
Matches 112; Conservative 39; Mismatches 72; Indels 40; Gaps 5;

QY 98 YQVYSLNFIYMIETFERHFDMLTKIHGSSFEKYPLYLAKVSGKEQTAKNAIWDG 157
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 8 YAKYNDMKNIVSMTEKMYEKHEPMVSRIRIGSTVEDNPLYLKI-GRKDGERRAIFMDCG 66

QY 158 IHAREWISPAFLMT------GHN-----
DB 67 IHAREWISPAFLMT------GHN-----RMR 180

QY 181 KRSFPAANHCIGTDLNRFNASKHMCCEGASSSEFYCGLYPESEPEKAVASFLRN 240
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 127 KRSNPNSTCTGTDLNRF-DVSWDSPTNDPCLSYVRGPAPESEKTKAVTNIRSH 185

QY 241 INQIKAYISMHSYSHIYFPPYTSKSDHEELISVASEAVRAIKTSKNTRYTHGHS 300
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 186 LMSIKAYIFHSYQMLLPYGYTKLPNHQDLKLVARIADVL-SSRYETRYIGPIA 244

QY 301 EFTYLAAGGGDDMYDLGIKYSF 323
DB 245 STIYKTSGLSDMAYDLGIKHTF 267

RESULT 9
CBP2_RAT STANDARD; PRT: 417 AA.
AC P19222;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN CPA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA MEDLINE=89034323; PubMed=3182871;
RA Gardell S.J., Craik C.S., Clausen E., Goldsmith E.J., Stewart C.-B.,
RA "A novel rat carboxypeptidase, CPA2: characterization, molecular
RT cloning, and evolutionary implications on substrate specificity in
RT the carboxypeptidase gene family."
RL J. Biol. Chem. 263:17828-17836(1988).
RN [2]
RP SEQUENCE OF 131-143 FROM N.A.
RX MEDLINE=95386501; PubMed=7657630;
RA Normant E., Gros C., Schwartz J.C.;
RA "Carboxypeptidase A isoforms produced by distinct genes or
RT alternative splicing in brain and other extrapancreatic tissues."
RL J. Biol. Chem. 270:20543-20549(1995).
RN [3]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=92105124; PubMed=1761558;

```

```

RA Fanning Z., Kobe B., Stewart C.-B., Rutter W.J., Goldsmith E.J.;
RT "Structural evolution of an enzyme specificity. The structure of rat
RT carboxypeptidase A2 at 1.9-A resolution."
RL J. Biol. Chem. 266:24606-24612(1991).
CC -1- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC residues.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M23721; AAA40956.1; JOINED.
DR EMBL: M23714; AAA40956.1; JOINED.
DR EMBL: M23715; AAA40956.1; JOINED.
DR EMBL: M23716; AAA40956.1; JOINED.
DR EMBL: M23717; AAA40956.1; JOINED.
DR EMBL: M23718; AAA40956.1; JOINED.
DR EMBL: M23719; AAA40956.1; JOINED.
DR EMBL: M23720; AAA40956.1; JOINED.
DR EMBL: M23721; AAA40956.1; JOINED.
DR PIR: A32128; A32128.
DR HSSP: P48052; IAYE.
DR MEROPS: M14.002; -.
DR InterPro: IPR003146; Proprep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Proprep_M14; 1.
DR Pfam: PF02246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CROXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 112
FT FT CHAIN 113 417
FT METAL 177 177
FT METAL 180 180
FT METAL 304 304
FT ACT_SITE 378 378
FT DISULFID 246 269
FT DISULFID 318 352
SEQUENCE 417 AA: 46912 MW: BEDCC41A830F2D45 CRC64;

Query Match 30.6%; Score 549.5; DB 1; Length 417;
Best Local Similarity 34.6%; Pred. No. 1.3e-39;
Matches 134; Conservative 63; Mismatches 127; Indels 63; Gaps 12;

QY 1 FQSGVYLAIPRTSRQYVQNLTTT--YEIVMOPYADLYKKQVHFVNASVDYV 58
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 20 FVGDVYLEIIPSHEDQIRTLQLEAEHELIDMWSKPT----IPGETVHVRVFPASIOAV 75

QY 59 KALINWSGIPCSYVLADVEDLIQOQISN--DIVSPASASYEYQHSLSNEIYSEFT 115
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 76 KYLESQGIDYSTIMIEDVYLDOEREMLFNORERGGFNEVYHTLEITIQEOMNLV 135

QY 116 ERHPDMLTKIHGSSFEKYPLYLAKVS-GKEQTAKNAIWDGIIHAREWISPAFLM-- 171
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 136 AENRGLVSKVNLGSSFEENRPMNLKSTGDD--KPAIMIDAGIHAREWVTOATALMTAN 192

QY 172 -----FT-----GH-----NMMKKNKSFYANNHCIGTDLN 197
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 193 KIASDYGTDPAITSLNTIDIFLLPYTNPDGYVFSQTTNMMKRTSKSGSCVGDPN 252

QY 198 RNFASKHMEGASSSSEFYCGLYPESEPEKAVASFLRNINQIKAYISMHSYSHI 257
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 253 RNM-DANFGPGASSSPCSDSYHGPRNSVEYKSIYDFIKSH-GYKAFITLHSTSQL 310

```

OY 258 VPPYSTRSKSDHEHLSIVASEAVRAIKRTSKNTRYTHGSEETLYLAPGGDDWYDL 317
 DB 311 MEPPGYKCYKPPDFNEHDEVAQKAQAL-KRLHGTSYKVGPCSVIYQASGSIDMAYDL 369
 OY 318 GIKYSTSN-----PVEKLLP 334
 DB 370 GIKISFAELRDTAFYGLLPKQILP 396
 RESULT 10
 CBPA_BOVIN STANDARD. PRT: 419 AA.
 ID CBPA_BOVIN STANDARD. PRT: 419 AA.
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase A precursor (EC 3.4.17.1).
 GN CPA.
 OS Bos taurus (bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID:9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-91151335; PubMed-1998496;
 RA Le Huecrou I., Guilloteau P., Toullec R., Puigserver A., Wicker C.;
 RT "Cloning and nucleotide sequence of a bovine pancreatic
 RT preprocarboxypeptidase A cDNA.";
 RL Biochem. Biophys. Res. Commun. 175:110-116(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA TISSUE-Pancreas;
 RA MEDLINE-96096549; PubMed-8522204;
 RA Goo J.H., Kim K.H., Choi K.Y.;
 RT "Cloning, sequencing and expression of the gene encoding a major
 RT allelic preprocarboxypeptidase A from bovine pancreas.";
 RL Gene 165:333-334(1995).
 RN [3]
 RP SEQUENCE OF 111-417.
 RA MEDLINE-71151183; PubMed-5102489;
 RA Bradshaw R.A., Walsh K.A., Neuraath H.;
 RT "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
 RT chymotryptic peptides of the cyanogen bromide fragment F-1.";
 RL Biochemistry 10:938-950(1971).
 RN [4]
 RP REVISIONS TO 138 AND 141.
 RA MEDLINE-72138789; PubMed-5143102;
 RA Petra P.H., Hermanson M.A., Walsh K.A., Neuraath H.;
 RT "Characterization of bovine carboxypeptidase A (Allan).";
 RL Biochemistry 10:4023-4025(1971).
 RN [5]
 RP SEQUENCE OF 17-120.
 RA MEDLINE-89150306; PubMed-3147705;
 RA Wade R.D., Hass G.M., Kumar S., Walsh K.A., Neuraath H.;
 RT "The amino acid sequence of the activation peptide of bovine pro-
 RT carboxypeptidase A.";
 RL Biochimie 70:1137-1142(1988).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
 RA MEDLINE-83294519; PubMed-6887246;
 RA Rees D.C., Lewis M., Lipscomb W.N.;
 RT "Refined crystal structure of carboxypeptidase A at 1.54-A
 RT resolution.";
 RL J. Mol. Biol. 168:367-387(1983).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
 RA MEDLINE-96003618; PubMed-7556081;
 RA Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
 RT "The three-dimensional structure of the native ternary complex of
 RT bovine pancreatic procarboxypeptidase A with proproteinase E and
 RT chymotrypsinogen C."

RL EMBL J. 14:4387-4394(1995).
 RN [8]
 RP VARIANT ALLELIC.
 RX MEDLINE-69283620; PubMed-5817619;
 RA Petra P.H., Bradshaw R.A., Walsh K.A., Neuraath H.;
 RT "Identification of the amino acid replacements characterizing the
 RT allelic forms of bovine carboxypeptidase A.";
 RL Biochemistry 8:2762-2768(1969).
 CC -1 CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O -> peptide + L-
 CC amino acid.
 CC -1 SUBUNIT: MONOMER. THE ZMOGEN IS SECRETED AS A TERNARY COMPLEX
 CC COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
 CC PROTEINASE E.
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -1 DATABASE: NAME-WORTHINGTON ENZYME MANUAL;
 CC WWW="http://www.worthington-biochem.com/manual/C/COA.html".
 CC
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M61851; AAA30426.1; -;
 CC EMBL: M61851; AAA30427.1; -;
 CC EMBL: 233906; CAA83955.1; -;
 CC PIR: JN0126; CEROA.
 CC PIR: A31406; A31406.
 CC PDB: 3CPA; 15-JAN-87.
 CC PDB: 4CPA; 22-OCT-84.
 CC PDB: 5CPA; 15-JAN-87.
 CC PDB: 6CPA; 15-OCT-91.
 CC PDB: 7CPA; 31-JAN-94.
 CC PDB: 8CPA; 31-JAN-94.
 CC PDB: 1CBX; 31-JAN-94.
 CC PDB: 1CPB; 15-OCT-94.
 CC PDB: 2CTB; 31-JAN-94.
 CC PDB: 2CTC; 31-JAN-94.
 CC PDB: 1ARL; 01-AUG-96.
 CC PDB: 1ARL; 17-AUG-96.
 CC PDB: 1BAV; 01-APR-97.
 CC PDB: 1YME; 12-FEB-97.
 CC PDB: 1CPX; 05-AUG-98.
 CC PDB: 1PYT; 27-JAN-97.
 CC MEROPS: M14.001; -;
 CC InterPro: IPR003146; Propep_M14.
 CC InterPro: IPR000834; Zn_carboxypept.
 CC Pfam: PF02244; Propep_M14; 1.
 CC Pfam: PF00246; Zn_carboxypept. 1.
 CC PRINTS: PR00765; CROBOXYPASA.
 CC PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 CC PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 CC Hydrolyase: Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
 CC 3D-structure; Polymorphism.
 CC SIGNAL 1 16
 CC PROPEP 17 110
 CC CHAIN 111 419
 CC METAL 179
 CC METAL 182 162
 CC METAL 306 306
 CC ACT_SITE 358 358
 CC ACT_SITE 380 380
 CC DISULFID 248 271
 CC VARIANT 289
 CC VARIANT 338
 CC VARIANT 415
 CC VARIANT 415
 CC CONFLICT 95
 CC CONFLICT 199
 CC CONFLICT 203
 CC CONFLICT 224
 CC
 CC I -> V (IN ALLELIC VARIANT).
 CC E -> A (IN ALLELIC VARIANT).
 CC L -> V (IN ALLELIC VARIANT).
 CC S -> L (IN REF. 5).
 CC D -> N (IN REF. 3).
 CC D -> N (IN REF. 3).
 CC D -> N (IN REF. 3).
 CC
 CC ACTIVATION PEPTIDE.
 CC CARBOXYPEPTIDASE A.
 CC ZINC.
 CC ZINC.
 CC ZINC.
 CC PROTON DONOR.
 CC NUCLEOPHILE.

FT CONFLICT 232 232 Q -> E (IN REF. 3).
 FT CONFLICT 295 295 D -> N (IN REF. 3).
 FT TURN 114 116
 FT TURN 119 120
 FT HELIX 125 138
 FT TURN 140 142
 FT STRAND 143 150
 FT TURN 152 153
 FT STRAND 156 162
 FT STRAND 171 176
 FT TURN 180 181
 FT HELIX 183 199
 FT TURN 200 202
 FT HELIX 204 212
 FT STRAND 214 218
 FT HELIX 223 231
 FT TURN 232 232
 FT TURN 234 235
 FT STRAND 242 242
 FT TURN 244 245
 FT STRAND 249 249
 FT TURN 253 254
 FT TURN 261 262
 FT STRAND 267 267
 FT TURN 270 271
 FT TURN 273 274
 FT STRAND 275 275
 FT TURN 280 281
 FT HELIX 284 296
 FT STRAND 299 307
 FT STRAND 311 314
 FT TURN 324 325
 FT HELIX 326 344
 FT STRAND 349 352
 FT HELIX 353 356
 FT TURN 357 357
 FT HELIX 364 370
 FT TURN 371 372
 FT STRAND 375 381
 FT TURN 388 389
 FT HELIX 393 416
 SEQUENCE 419 AA; 47082 MM; 21B86407B3BFC452 CRC64;

Query Match 29.8%; Score 536.5; DB 1; Length 419;
 Best Local Similarity 34.8%; Pred. No. 1.6e-38;
 Matches 129; Conservative 59; Mismatches 124; Indels 59; Gaps 10;

QY 1 FOSGOVLAALPRTSKOVYLOLTTTTEIVL--W---QPTADLIYKKQVHEFYNASD 54
 DB 20 FVGHQVLRITTADEAEVQTKLEDELEHQLDPMRGPGQP-----GSPIDVRVPFPS 71
 QY 55 VONVKAHLNVSGIPCSVLADVEDLI---QOQISNDTVSPRASASY-YEOYSLNLTYSM 110
 DB 72 LQAVAVFLFAHRIKIRIMIEDVQSLDEQEQMFASQSAKSTNFTFNATVHTLDTIYFP 131
 QY 111 EEFITERHEPDMLTIKIHIGSSFEKYPYLVKVGKEQTAKNAIWDGIAHREWISPAFL 170
 DB 132 MOLVAHEHQVLQSLQIGSYGRPIYLVKFS-TGSSNRPALIMIDGISHREMITQATGV 190
 QY 171 WFI-----GNNRMKRRKRSFYANNHCIG 193
 DB 191 WFAKFTEDYGDPSFTALIDSMDFLEIVTNPDPGFAPTHSONRLMRKRTRSVTSSSLCYG 250
 QY 194 TOLNRFASKHMCESGSSSEYCYGLYPESEPEVKAVASFLRNINQIKAYISMHSY 253
 DB 251 VVANRRW-DAGGKACASSPCSEYTHGKXANSEVVKSIYDFVKDGN-FKAFLSHSY 308
 QY 254 SCHIYFPYSYTRSKSDHELSLVASEAVRAIKTSKNTRYTHGSGFTLYLAPGGGDM 313
 DB 309 SOLLLYPYCITQSIQIPDKELNOVAKSAVEAL-KSLYGSYKKGSITITTYQASGSDIM 367
 QY 314 IYDLGIRKYSFT 324

DB 368 SYNGIKYSFT 378
 RESULT 11
 ID CBP1_RAT STANDARD; PRT; 419 AA.
 AC P00731;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
 GN CPAL.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP MEDLINE-82105986; PubMed-6275388;
 RX Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standing D.N.,
 RA Picot R.L., Valenzuela P., Rutter W.J.;
 RT "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
 RT characterization of the gene." Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-89034324; PubMed-3182872;
 RA Clausen E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
 RT "Structural characterization of the rat carboxypeptidase A1 and B
 RT genes. Comparative analysis of the rat carboxypeptidase gene
 RT family." J. Biol. Chem. 263:17837-17845(1988).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
 CC amino acid.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14, ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: V01232; CAA24542.1; -;
 DR EMBL: J00713; AAA40893.1; -;
 DR EMBL: M23990; AAA40955.1; -;
 DR EMBL: M23960; AAA40955.1; JOINED.
 DR EMBL: M23985; AAA40955.1; JOINED.
 DR EMBL: M23986; AAA40955.1; JOINED.
 DR EMBL: M23987; AAA40955.1; JOINED.
 DR EMBL: M23988; AAA40955.1; JOINED.
 DR EMBL: M23989; AAA40955.1; JOINED.
 DR PIR: A00911; CPTA.
 DR PIR: B32129; B32129.
 DR HSSP: P00730; 1PPT.
 DR MEROPS: M14.001;
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxypept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxypept. 1.
 DR PRINTS: PR00765; CRBOXYPRASRA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolyase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
 FT SIGNAL 1 16
 FT PROPEP 17 110 ACTIVATION PEPTIDE.
 FT CHAIN 111 419 CARBOXYPEPTIDASE A1.
 FT METAL 179 179 ZINC (BY SIMILARITY).
 FT METAL 182 182 ZINC (BY SIMILARITY).


```

DR Interpro: IPR003146; Propep_M14.
DR Interpro: IPR000834; Zn_carboxypept.
DR Pfam: PF00244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept; 1.
DR PRINTS: PR00765; CRBOXYPTASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 113
FT CHAIN 114 421
FT METAL 181 181
FT METAL 184 184
FT METAL 308 308
FT ACT_SITE 382 382
FT DISULFID 250 273
SQ SEQUENCE 421 AA; 47379 MW; 91883D25B705B6A CRC64;

Query Match 27.3%; Score 491; DB 1; Length 421;
Best Local Similarity 32.7%; Pred. No. 1.3e-34;
Matches 123; Conservative 59; Mismatches 126; Indels 68; Gaps 11;

QY 1 FGGGVLAALPRTSRQVQLQNLTTTEIYL--WQVTA-----DLIVKKKQVHEFNAS 53
DB 21 FFGQVLRIRVNGDELSKLSQLYNSNNLKNLKNFKSSFRNPVAVLPVSYLQAF---- 76
QY 54 DVDNKAHLNVSGIPCSVLLADVEDLI--QQQISNDTVSPRASASY-YEQYHSLNETYS 109
DB 77 -----KSELSQGLEAVVTIEDLQALLDNDEDEKQHNDEGERSNNNYGAYHSLDAIYH 131
QY 110 WIEFTRHNDMLTKIHGSEFEKYPYLVKVSKEQTAANKATIDCGIAREWISAPFC 169
DB 132 EMDNIAADPDLARVKGICHSFENRPMYVLKFSYKGVRRPAVVLNGLHSREVISQATA 191
QY 170 LW-----FI-----GHRMRKNRSFFANNHCI 192
DB 192 IWRARKVSYQDRPATSLLEKMDIFLLPVANDGYVYOTQNRMLKRTSRKRPSSCI 251
QY 193 GTDLNRN---PASKHWCCEGASSSCSETEYCGLYPSEPEPKAVASFLRRNINOIKAYI 248
DB 252 GADPNRMANNSAFGR-----GASDNPCESEVYHGHANSEVEKSYVDIQLHGN-FKGI 305
QY 249 SMHSYSGHIVFPYSYTRSKSKDHEELSLVASEAVRALEKTSKNTRYTHGHGSETLY 308
DB 306 DLHSYSQLLMYPGYGYKAKDADELDKVAKLAKALASVS-GTEYOVGPTCTTVYPSG 364
QY 309 GGDDWIYDLGIRKYSFT 324
DB 365 SSIDMAYDNGIKFAFT 380

RESULT 15
CBPZ_SIMVI STANDARD; PRT; 304 AA.
AC P42788;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
OS Simulium vittatum (Black fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Simuliidae; Simulium.
OX NCBI_TaxID=7192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=gut;
RX MEDLINE=9403864; PubMed=8269093;
RA Ramos A., Mahowald A., Jacobs-Lorena M.;
RT "Gut-specific genes from the black fly Simulium vittatum encoding
RT trypsin-like and carboxypeptidase-like proteins.";
RL Insect Mol. Biol. 1:149-163(1993).

```

```

CC -1- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
CC -1- TISSUE SPECIFICITY: GUT-SPECIFIC.
CC ZINC SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
CC EMBL: L08481; AAA18531.1; -.
CC HSSP: P48052; IAYE.
DR Interpro: IPR000834; Zn_carboxypept.
DR Pfam: PF00246; Zn_carboxypept; 1.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc.
FT NON_TER 1 1
FT METAL 58 58
FT METAL 61 61
FT METAL 184 184
FT ACT_SITE 236 236
FT ACT_SITE 259 259
FT DISULFID 125 148
SQ SEQUENCE 304 AA; 34849 MW; 2E6E3F8A6A9144 CRC64;

Query Match 22.9%; Score 411.5; DB 1; Length 304;
Best Local Similarity 33.3%; Pred. No. 5.3e-28;
Matches 94; Conservative 52; Mismatches 85; Indels 51; Gaps 9;

QY 100 QYHSLNEIYSWIEFTRHNDMLTKIHGSEFEKYPYLVKVSKEQTAANKATIDCGI 159
DB 1 QYHSLNEIYSWIEFTRHNDMLTKIHGSEFEKYPYLVKVSKEQTAANKATIDCGI 159
QY 160 AREWISPA---FCL-----WFI-----GH-----NRMRKNRS 184
DB 59 AREWITATATYTYLLNELTSKNSITIREMAENVDYITPVPNDGYVYTHHTDRMRKRTS 118
QY 185 FYANHCIGTDLNRNPAKSHWCCEGASSSCSETEYCGLYPSEPEPKAVASFLRRNINOI 244
DB 119 PNPISLCACTDPNRNW-NFHMEDGTSRPTCTEYGGKAFSEVETNSFSFLTLGQI 177
QY 245 KAYISMSYSGHIVFPYSYTRSKSKDHEELSLVASEAVRALEKTSKNTRYTHGHGSETLY 304
DB 178 KVLAFHSYSQLLFPYGHGTCQHTYHNDLQALGDAARSLAQ-RYGDYTVGNITDAIY 236
QY 305 LAPGGDDWIY---DLGIRKYSFTSNP-----PYEKLP 334
DB 237 PASGGSDMAYDLDIPIAYTYELRPDGMWNGFOLPANQITP 278

```

Search completed: September 18, 2002, 16:47:50
Job time: 282 sec

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnVal 17
 84 TTCACAGAGTGGCCAGTTCTAGCTGCTTCCACGAACCTCTAGCCAGT 133
 17 LglnValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTyrGlnP 34
 134 TCAGAGTCTACAGAAATCTTACTACAAACATATGATGATGTTCTCTGCAG 183
 34 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
 184 CGGTACAGCTGACCTTATTGTGTAAGAAAAACAACTCCATTTTGTGTA 233
 51 AsnAlaSerAspValAlaAsnValLysAlaHisLeuAsnValSerGly11 67
 234 AATGCATCTGATGTCGACAAATGCAAGCCCAATTTAAATGTGAGCGGAAT 283
 67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln1 84
 284 TCCATCAGTGTCTTGTGTCGACAGCTGGAGATCTTATTACACACAGA 333
 84 LeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGln1 100
 334 TTTCCAGACACACAGACAGCCCGGAGCTCCGCACTGACTATGACAG 383
 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluAr 117
 384 TATCACTACTAAATCAATCTATCTGTGATGAAATTTATTAACAGAG 433
 117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluTyr 134
 434 GCATCTGATGATGCTTACAAAATCCACATGTGATCTCTATTGAGAAAT 483
 134 TyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
 484 ACCCACTATGTTTAAAGTTTCTGGAAAGAAACAAACACCAAAAT 533
 151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
 534 GCCATATGATGATGCTGGAATCCATGCCAGAGATGATCTCTCCCTCC 583
 167 aPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArgLysAsnArgS 184
 584 TTTCTGCTTGTGTCATAGGCCATATCGAATGTGGAGAAAGAACCTGT 633
 184 ePheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
 634 CTTTCTATGCGAAACATCATTCATCGAAGACACCTGAAATGACAACTTT 683
 201 AlaSerLysHisTrpCysGluGlnValLysSerSerSerCysSerG1 217
 684 GTCCTCAAAACACTGCTGTGAGGAAAGTCATCCAGTCCCTCACTGCTCGA 733
 217 uThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValA 234
 734 AACCTACTGTGACTTATCTCGAGTCAGAAACCAAGTGAAGGACAGTGG 783
 234 LaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMet 250
 784 TAGAGTTCTTGAAGAAATATCAACCCAGATTAAAGCATACATCAGCATG 833
 251 HisSerTyrSerGlnHisIleValPheProTyrSerTyrTrpArgSerLy 267
 834 CATTCATCTCTCCAGCAATATAGTGTTCATATTCCTATACACGAAGTAA 883
 267 sSerLysAspHisGluGlnLeuSerLeuValAlaSerGluAlaValArgA 284
 884 AAGCAAGACCATGAGAACTGTCTTAGTAGCCAGTCAACAGACTGCTGG 933
 284 LaIleGluLysThrSerLysAsnThrArgTyrTrpHisGlyHisGlySer 300
 934 CTATTGACAAACATAGTAAATAACAGGTATACACATGGCCATGGCTCA 983

301 GluThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLe 317
 984 GAAACCTTATACCTAGCTCTGGAGGTGGGACAGATTCATCTATGATTT 1033
 317 uGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuP 334
 1034 GGGCATCAAAATATGCTTACATCAAAACCCACTGTAGAGAAAGCTTTTGC 1083
 334 roLeuSerLeuLys 338
 1084 CCCTGCTCTGTAATAA 1097

seq_name: /SID51/5cdata/geneseq/geneseqn-emb1/NA1999.DAT.AAV74302

seq_documentation_block:
 ID AAV74302 standard; cDNA; 1272 BP.

AAV74302;

28-APR-1999 (first entry)

Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.

Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;

polymorphism detection; thrombotic disease; ds.

Homo sapiens.

W09855645-A1.

10-DEC-1998.

02-JUN-1998; 98WO-EP03244.

03-JUN-1997; 97US-0869057.

(SCHD) SCHERING AG.

Morser MJ, Nagashima M;

WPI: 1999-045800/04.

P-PSDB; AAM92270.

Detecting new polymorphism of human plasma carboxypeptidase B -
 PT comprises Alanine or Threonine at position 147 of protein by DNA or
 PT protein analysis, useful to detect risk of thrombotic disease in
 humans

Example 1; Page 24; 35pp; English.

This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
 CC hPCPBthr147. The invention relates to a method for determining the
 CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
 CC comprises obtaining a prepared tissue or blood sample and determining the
 CC presence of DNA coding for naturally occurring polymorphisms of the protein
 CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
 CC respectively). Determination of the relative distribution of the PCPB
 CC polymorphs in a patient's blood by genetic or protein analysis by the
 CC methods is useful to determine the risk of thrombotic disease in humans.
 CC Such assessments may be made by accumulating information concerning the
 CC relative distribution of the different polymorphs within the general
 CC population compared with populations known to be at risk and establishing
 CC a PCPB polymorph profile for at-risk patients.

Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

alignment_scores:

Quality: 1732.00 Length: 393
 Ratio: 5.124 Gaps: 3
 Percent Similarity: 86.005 Percent Identity: 85.751

alignment_block:

US-09-980-881-3 x AAV74302 ..

Align seg 1/1 to: AAV74302 from: 1 to: 1272

```

1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnVal 17
  |||
67 TTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTTGAACCTCTAGGCAAGT 116
  |||
17 LglnValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGln 34
  |||
117 TCAAGTCTTCAGAAATCTTACTACAACTATGATGTTCTCTCGGAGC 166
  |||
34 roValThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
  |||
167 CGGTAAACAGTGAACCTTATGTGAGAAAAAACAGCCATTTTGTGTA 216
  |||
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyI 67
  |||
217 AATGCATCTGATGTCCGAAATGTGAAGCCCATTTAAATGTGAGCGGAA 266
  |||
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
  |||
267 TCCATGCAGTGTCTGTGCTGGCAGACGTGGAAGATCTTATTCACAGCAG 316
  |||
84 IeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGln 100
  |||
317 TTTCCACAGCACAGTCAAGCCCGGAGCCCTCCGATCTACTATGAGCAG 366
  |||
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 117
  |||
367 TATCACTCACTAAATGAAATCTTCTTGGATGAAATTTATTAAGTAGAG 416
  |||
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
  |||
417 GCATGCTGATATGCTTACAAATAATCCACATTTGATCTCATTTGAGAG 466
  |||
134 yTrpLeuTyrValLeuLysValSerGlyLysGluIleThrAlaLysAsn 150
  |||
467 ACCCACTCTATGTTTAAAGTTCTGGAAGAAACAAAGCCAAAAT 516
  |||
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerPro 167
  |||
517 GCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 566
  |||
167 aPheCysLeuTrpPheIleGlyHis..... 175
  |||
567 TTTCTGCTGTGTTGATCATAGCCATATACTCAATTCATGGATATAG 616
  |||
175 ..... 175
  |||
617 GGCAATATACCAATCTCTGAGGCTTGTGATTTCTATGTTATGCCGTC 666
  |||
176 ..... AsnArgMetTrp 180
  |||
667 GTTAATGTGACGTTATGACTCTCATGGAATAAGATGATGAG 716
  |||
180 gLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeu 197
  |||
717 AAGAACCGTTCTTCTATGCGAAACATATGATGATGATGATGATGATG 766
  |||
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 213
  |||
767 ATAGGAACCTTGTCTTCCAAACACTGTGTGAGGAAGTGCATCCAGTTCC 816
  |||
214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGlu 230
  |||
817 TCATGCTCGAAACCTTACTGTGACTTATCTGAGTCAACACAGCAAGT 866
  |||
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGluIleLysAla 247
  |||
867 GAAGCAGGTGGCTAGTTCTTGAAGAAATATCAACACAGATTAAGCAT 916
  |||
247 yTrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263

```

```

|||||
917 ACATCAGCATGCAATTCATCTACCCAGCATATAGTGTTCATATCTCTAT 966
  |||
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
  |||
967 ACACGAAGTAAAGCAAGACATGAGGAACTGTCTCTAGTACGAGTGA 1016
  |||
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHis 297
  |||
1017 AGCAGTTCGTGCTATTGAGAAACTAGTAAATAATACCAAGTATACATG 1066
  |||
297 LysIleGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrp 313
  |||
1067 GCCATGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGAGCATGG 1116
  |||
314 IleTyrAspLeuGlyIleLysTyrSer.Phe..... 323
  |||
1117 ATCTATGATTTGGCATTCAATATTCGTTTCAATTTGAACTTCAGATAC 1166
  |||
324 .....ThrSerAsnProProVal 329
  |||
1167 GGGCAGATAGGATTTCTTGTGCGCGGAGCGTTACATCAACACCGTGA 1216
  |||
330 GluLysLeuLeuProLeuSerLeuLys 338
  |||
1217 GAGAAAGCTTTTGGCCGTGTCTCTTAAA 1243
  |||
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA196.DAT:AA11671
seq_documentation_block:
ID AA11671 standard; DNA; 1749 BP.
XX
AC AA11671;
XX
DT 12-APR-1996 (first entry)
XX
DE Human plasma carboxypeptidase B coding sequence.
XX
KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 41..1312
FT /tag= a
FT /product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT /tag= b
FT mat_peptide 107..1309
FT /tag= c
XX
PN US5474901-A.
XX
PD 12-DEC-1995.
XX
XX
PF 01-FEB-1991; 91US-0649591.
XX
PR 01-FEB-1991; 91US-0649591.
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
XX
PA (GENE ) GENENTECH INC.
XX
PI Drayna DT, Eaton DL;
XX
XX WPI; 1996-039508/04.
XX DR P-PSDB; AAR90293.
XX
PT Antibody to human plasma carboxypeptidase B - useful for detecting
PT and purifying hPCPB for use in treating clotting disorders e.g.
PT haemophilia A

```

XX Disclosure: Figure 4: 40pp: English.
 PS
 CC An antibody which specifically binds human plasma carboxypeptidase B
 CC (hPCPB) and does not cross react with other carboxypeptidases is
 CC useful for the detection of hPCPB in vitro. The antibody is also
 CC used for purifying hPCPB from a sample. Purification comprises
 CC passing a sample thought to contain hPCPB over either a column to
 CC which antibody has been bound, or a plasminogen affinity column,
 CC eluting the column and then recovering the fraction containing the
 CC hPCPB.
 CC
 SQ Sequence 1749 BP: 521 A; 361 C; 342 G; 525 T; 0 other:

 Alignment_scores:
 Quality: 1732.00 Length: 393
 Ratio: 5.124 Gaps: 3
 Percent Similarity: 86.005 Percent Identity: 85.751

 Alignment_block:
 US-09-980-881-3 x AAT11671 ..
 Align seg 1/1 to: AAT11671 from: 1 to: 1749

 1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgFhSerArgGlnVa 17
 107 TTTCAAGAGTGGCCAAAGTCTAGCTGCTCTCTAGAACCTTAGGCAAGT 156
 17 LcInValLeuGlnAsnLeuThrThrThyTyrGluLeuValLeuTyrPGLP 34
 157 TCAAGTCTACAGAACTTACTACAAACATATGAGATGTTCTCTGCGACC 206
 34 roValThrAlaAspLeuLeuValLysLysLysGlnValHisPhePheVal 50
 207 CGCTAACACCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTGTGA 256
 51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyL 67
 257 AATGATCTGATGTGACAAATGTGAAAGCCATTAAATGATGAGCGGAAAT 306
 67 eProCysSerValLeuLeuAlaAspValGluAspLeuLeuGlnGlnLni 84
 307 TCCATGCACTGCTGCTGCGAGACGTGGAAGATCTATTATCAACAGCAGA 356
 84 LeSerAsnAspThrValSerProAlaSerAlaSerTyrTyrGluGln 100
 357 TTTCCAAGCAGACAGTCAAGCCCCCGAGCCTCCGATCTGACTATGAAACAG 406
 101 TyrHisSerLeuAsnGluLeuTyrSerTyrPheGluPheLeuThrGluAr 117
 407 TATCACTCACTAAATGAAATCTATCTTGATAGAAATTTATATACAGAG 456
 117 GHisProAspMetLeuThrLysLeuHisLeuGlySerSerPheGluLysT 134
 457 GCATCTCATATGCTTACAAAAATCCACATTCGATCCATTTGAGAGT 506
 134 yrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
 507 ACCCCTCTATGTTTAAAGTTTCTGGAAGAAACAAACCCCAAAAT 556
 151 AlaIleTyrPheAspCysGlyLLeuHisAlaArgGluTyrPheSerProAl 167
 557 GCCATATGATGATGACTGTGAAATCCATGCCAGAAATGATCTCTCTGCC 606
 167 aPheCysLeuTyrPheLeuGlyHis..... 175
 607 TTTGCTGTTGTTGATCAGCCATATACATCAATTTATATGAGATATAG 656
 175 175
 657 GGCATATATACCAATCTCTGAGGCTTGTGATTTCTATATATAGCCGCTG 706

176AsnArgMetTyrPar 180
 707 GTTAATGTGAGCGTTATGACTACTATCATGTGAAAAAGAAATCGAAATGTGGAG 756
 180 GlyAsnAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuA 197
 757 AAGAAGCCCTTCTTCTATGCGAAACATCATTCATCGGAAACAGACCTGA 806
 197 snArgAsnPheAlaSerLysHisTyrCysGluGluGlyLysSerSer 213
 807 ATTAGAACTTGTCTTCCAAACACTGTGTGAGAGAGGTGCATCCAGTTCC 856
 214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVa 230
 857 TCATGCTCGAAACCTACTGTGACTTATCTCTGAGTCAAGAACAGAGT 906
 230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaT 247
 907 GAAGCGAGTGGCTAGTTCTTGAGAAATATCAACACGATTAAGCAT 956
 247 yrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
 957 ACATGACATGCACTCATCTATCTCCAGCATATATAGTGTTCATATTCCTAT 1006
 264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
 1007 ACACGAAGTAAAGCAAGACCATGAGGAACGTCTCTAGTAGCCAGTGA 1056
 280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisG 297
 1057 AGCAGTCTGCTATTGAGAAACTAGTAAATATCAAGATATACACATG 1106
 297 LysHisGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAspAspTyrP 313
 1107 GCCATGGCTCAGAAACCTTATACCTAGCTCTCGAGAGTGGGAGCATTCG 1156
 314 IleTyrAspLeuGlyIleLysTyrSer.Phe..... 323
 1157 ATCTATGATTTGGGCATCAAAATATCGTTTACAAATGAACTTCGAGATAC 1206
 324ThSerAsnProProVal 329
 1207 GGGCACAATACGATTTCTTGCGCGAGCGATTACATCAAAACCACTGTGA 1256
 330 GluLysLeuLeuProLeuSerLeuLys 338
 1257 GAGAAGCTTTGGCCGCTGCTCTTAA 1283

 seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62846
 seq_documentation_block:
 ID AAT62846 standard; DNA: 1749 BP.
 XX AAT62846;
 AC
 XX
 DT 08-MAY-1997 (first entry)
 DE Human plasma carboxypeptidase B coding sequence.
 XX
 KW Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
 XX plasma; plasminogen; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 41..1312 /tag= a
 FT FT /product= Human PCPB
 FT sig_peptide 41..106 /tag= b
 FT mat_peptide 107..1309 /tag= c
 FT FT
 FT misc_binding 134..177

```

FT      /*tag= d
XX      /bound_moelty= 46_bp_probe
XX      US5593674-A.
XX      14-JAN-1997.
XX      01-FEB-1991; 91US-0649591.
XX      01-FEB-1991; 91US-0649591.
XX      01-FEB-1991; 91US-0649591.
XX      14-OCT-1992; 92US-0959944.
XX      15-DEC-1993; 93US-0167727.
XX      19-JUL-1994; 94US-0277540.
XX      27-APR-1995; 95US-0430787.
XX      (GETH ) GENENTECH INC.
XX      Drayna DF, Eaton DL;
XX      WPI: 1997-099413/09.
XX      P-PSDB; AAM14733.
XX      Using human plasma carboxypeptidase B in blood coagulation - is
XX      functionally related to carboxypeptidase A and pancreas
XX      carboxypeptidase B
XX      Example 2; Column 37-42; 39pp; English.
XX      This sequence encodes human plasma carboxypeptidase B (PCPB) which
XX      has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
XX      PCPB may be used therapeutically in haemostatic regulation. PCPB is
XX      purified from human plasma or by transformed cell culture by
XX      extraction using plasminogen bound to a solid phase.
XX      Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other:

alignment_scores:
    Quality: 1732.00      Length: 393
    Ratio: 5.124          Gaps: 3
    Percent Similarity: 86.005      Percent Identity: 85.751

alignment_block:
US-09-980-881-3 x AAT62846 ..

Align seg 1/1 to: AAT62846 from: 1 to: 1749

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnVal 17
107 TTTCAGAGTGGCCAAAGTCTAGCTGCTCTTCTAGAACCTTAGGCNAAGT 156
17 LcGlnValLeuGlnAsnLeuThrThrThrTyrgluIleValLeuTyrglnP 34
157 TCAAGTTTACAGAAATCTTACTACACATATGATGATGTTCTCTGGCAGC 206
34 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
207 CGGTAAACACCTGACCTTATTTGCAAGAAAAACAAGTCCATTTTGTGTA 256
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyI 67
257 AATGCATCTGATGTCACAAATGTGAAAGCCCATTTAAATGTGAGCGAAT 306
67 eProCysSerValIleuLeuAlaAspValGluAspLeuIleGlnGlnGln 84
307 TCCATGCACTGCTCTGCTGCGAGAGCTGGAAGATCTTATTCACACAGAG 356
84 lSerAsnAspThrValSerProArgAlaSerAlaSerTyrgluGln 100
357 TTTCACACACACAGTACAGCCCCGAGCCCTCCGATCTACTATGAAACAG 406
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluAr 117
|||||

```

```

407 TATCACTCACTAAATGAATCTATTTCTGATGATGCAATTTTAACTGACAG 456
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTr 134
457 GCATCCCTGATATGCTTACAAAAATCCACATTTGGATCTCTCATTTTGAGAGT 506
134 yProLeuTyrValLeuLysValSerGlyLysGluIleThrAlaLysAsn 150
507 ACCCACTCTATGTTTAAAGTTTCTGAAAAAGAACAAACAGCCCAAAAAAT 556
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
557 GCCATATGATTTACTGTGATGATCCATGACAGAAATGATCTCTCTGC 606
167 apheCysLeuTrpPheIleGlnHis..... 175
607 TTTCGCTGTGTGTCATATAGCCCATATATCAATCTATGATGATATAG 656
175 ..... 175
657 GGCAATATACCAATCTCCCTGAGGCTTGTGATTTCTATGTCGCCGCTG 706
176 ..... AsnArgMetTrpAr 180
707 GTTAATGTGACGCGTTATGACTCTCATGCAAAAAGAAATGATGAG 756
180 gLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuA 197
757 AAGAACCCTGCTTTCTATGCGAACAATCATTTGCTGGAACAGACCTGA 806
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 213
807 ATAGGAACCTTGCTCCAAACACTGTGTGAGGAGAGCTGATCCAGTCTCC 856
214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVa 230
857 TCATGCTCGAAGACCTACTGTGGCTTATCTGATCTGACAGACAGAGT 906
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaT 247
907 GAAGGACGTGGTAGTTCTTGAGAAAGAAATATCAACCATTAATTAACAT 956
247 yTrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
957 ACATACACATGATCATCTACTCCAGCATATAGTGTTCATATCTCTAT 1006
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
1007 ACACGAAAGTAAAGCAAGACATGAGAACTGTCTCTAGTAGCAGGTGA 1056
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisG 297
1057 AGCAGTTCGTGCTATGTAGAAAACTAGTAAAAATACAGATATACACATG 1106
297 LysHisGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAspAspTrp 313
1107 GCCATGCTCTAGAAACCTTATCTACTCTCTGAGGTGGGAGAGTATGG 1156
314 lIleTyrAspLeuGlyIleLysTyrSer.Phe..... 323
1157 ATCTATGATTTGGGATCAAAATATTCGTTTCAACATTAACCTGAGATAC 1206
324 ..... ThrSerAsnProProVal 329
1207 GGGCACATACGAGATTTCTGCTGGCGAGCGTTACATCAAAACCCAGCTGA 1256
330 GluLysLeuLeuProLeuSerLeuLys 338
1257 GAGAAAGCTTTTGGCGGCTGTCTTAAAA 1283
seq_name: /SIDSI/gcgsdata/geneseq/geneseqn-emb1/NA1993.DAT:AA041001
seq_documentation_block:

```

ID AA041001 standard; cDNA; 1749 BP.
XX
AC AA041001;
XX
DT 24-AUG-1993 (first entry)
XX
DE Human plasma carboxypeptidase B gene.
XX
KW PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX
OS Homo sapiens.
XX
FH key Location/Qualifiers
FT sig_peptide 41..106
FT /*lag- a
FT 107..1312
FT /*lag- b
FT /*product- PCPB
FT 134..177
FT misc_feature
FT /*lag- c
FT /*function- probe
FT /*note- "used to obtain full-length clones"
XX
PN US5206161-A.
XX
PD 27-APR-1993.
XX
PE 01-FEB-1991; 91US-0649591.
XX
PR 01-FEB-1991; 91US-0649591.
XX
PA (GETH) GENENTECH INC.
XX
PI Drayna DT, Eaton DL;
XX
DR WPI; 1993-151724/18.
XX
DR P-PSDB; AAR36273.
XX
PT New human plasma carboxypeptidase B - used as haemostatic
PT regulator for clotting blood, partic. for treating blood clotting
PT disorders, e.g. haemophilia
XX
PS Disclosure; Fig 4; 40pp; English.
XX
CC Human plasma carboxypeptidase B was isolated from human plasma and
CC partially sequenced. Oligonucleotide primers were designed based on
CC the partial amino acid sequences. The primers were used in a PCR
CC amplification to identify cDNA encoding PCPB from a human liver cDNA
CC library. The PCR product was capable of encoding the first 37 amino
CC acids of PCPB; a 46mer probe was used to obtain the full-length
CC sequence which, although disclosed in the specification, is not
CC claimed. PCPB inhibits the enzymatic conversion by tPA of
CC plasminogen to plasmin in the presence of fibrinogen.
XX
SQ Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

alignment_scores:
Quality: 1727.00 Length: 393
Ratio: 5.125 Gaps: 3
Percent Similarity: 85.751 Percent Identity: 85.496

alignment_block:
US-09-980-881-3 x AA041001 ..

Align seg 1/1 to: AA041001 from: 1 to: 1749

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnVal 17
|||||
107 TTTCAGAGTGGCCAGATTCTAGCTGCTCTTCTAGAACCTCTTAGGCAAGT 156
|||||
17 GlnValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTyrPlnP 34
|||||

157 TCAGTTCCTACAGAAATCTTACTACAAACATATGAGATTGTTCTCTGCGACG 206
34 rovalThrAlaAspleuIleValLysLysGlnValHisPhePheVal 50
|||||
207 CGGTACAGCTGACCTTATGTGAAAGAAAAAACAGATCTTTTGTGTA 256
51 AsnAlaSerAspValAspAsnValLysAlaHisIleAsnValSerGlyTl 67
|||||
257 ATATGATCTGATGTCGACAAATGTGAAAGCCCATTTAAATGTGAGCGAAT 306
67 eProCysSerValLeuAlaAspValGluAspleuIleGlnGlnI 84
|||||
307 TCCATGACAGTGTCTGTGGCAGACGTGAAAGATCTTATTCACAGCAGA 356
84 IeSerAsnAspThrValSerProAlaGlnAlaSerTyrTyrGluGln 100
|||||
357 TTCCACAGCAGACAGTACGCCGCCGAGCTCCGATCTCTACTATGACAG 406
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluAr 117
|||||
407 TATCCTCAGTAAATGAATCTATTTCTGATAGATTTATTAAGTGAAG 456
117 GHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluTyr 134
|||||
457 GCATCCTGATATGCTTACAAAAATCCACATTCGATTCCTCATTTGAGAAGT 506
134 YrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
|||||
507 ACCACCTATATGTTTAAAGTTTCTGAAAGAAAGAAACCAAGCCCAAAAT 556
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
557 GCCATATGATGATGACTGTGAATCCATGCCAGAAATGATGATCTCTGTC 606
167 aPheCysLeuTrpPheIleGlyHis..... 175
607 TTTCGTTGTGTTCATAGGCGCATATTAATCAATCTATAGGATTAATAG 656
175 175
657 GGCATATACCAATCTCTGAGGCTTGTGATTTCTATGTTATCCCGGTG 706
176AsnArgMetTrpPar 180
707 GTTATGTGACGCGTTATGACTACTCATGAAAAAGATGAATGTGGAG 756
180 GlyAsnAspSerPheTyrAlaAsnAsnHisCysIleGlyThrAspleuA 197
757 AAAGAACCGTTCTTCTATGCGAACAATCATGTCATGGAACAGACCTGA 806
197 snArgAsnPheAlaSerLysHisIleTrpCysGluGluGlyAlaSerSerSer 213
807 ATAGAACTTGTCTCCAAACACGTGAGTGAAGATGATGACAGTCC 856
214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVal 230
857 TCATGCTCGGAAACCTACTGTGACTTATCTGAGTCAAGAACACAGAGT 906
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGluIleLysAlaLe 247
907 GAAGCGAGTGGCTAGTTCTTGGAGAAATATCAACCAATTAAGCAT 956
247 YrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
957 ACATACAGCATGATTCATCTCCAGCATATAGTTCATATTCCTAT 1006
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
1007 ACACGAAGTAAAGCAAAACATGAGGAACCTCTCTAGTACCGCAGTGA 1056
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisG 297
1057 ACAGCTTCTGTCTATTCAGAAAACTAGTAAAAATACAGGATATACACATG 1106

297 LNHISGLYSERGLUTRHLTYRLEUALPROGLYGLYASPAAPTTP 313
 1107 GCCATGGCTCAGAAACCTTATACCTAGCTCTGAGAGTGGGAGCATTTGG 1156
 314 LLeTYRAsPLeuGLYLeuTYRSeR.Phe..... 333
 1157 ACTATGATTTGGGCATCAAAATATGTTACAAATGAACTTGAGATAC 1206
 324ThSeRAsnProPoval 329
 1207 GGGCACAATACGATTTCTGCTGCCGAGCGCTTACATCAAAACCACTGTA 1256
 330 GLuLYSLeuLeuProLeuSeRLeuLYs 338
 1257 GAGAACCTTTTGGCCGCTGCTCTAATAA 1283
 seq_name: /SIDSL/9cdata/geneseq/geneseqn-emb1/NA2000.DAT:AAF18005
 seq_documentation_block:
 ID AAF18005 standard; DNA; 1400 BP.
 AC AAF18005;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Lung cancer associated polynucleotide sequence SEQ ID 24.
 XX
 KW Human; lung cancer associated protein; neuroprotective; cytostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotoxic; antiinfective; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 OS Homo sapiens.
 XX
 PN WO200055180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Ruben SM;
 XX
 DR WPI: 2000-587514/55.
 DR P-PSDB: AAB58129.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 1; Page 507; 1425bp; English.
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytoskeletal; cardioactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
 CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and

CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;
 alignment_scores:
 Quality: 902.00 Length: 238
 Ratio: 4.983 Gaps: 4
 Percent Similarity: 76.050 Percent Identity: 75.210
 alignment_block:
 US-09-980-881-3 x AAF18005 ..
 Align seg 1/1 to: AAF18005 from: 1 to: 1400
 157 GLYLeuHISAlaRgGLUTRPLLeSeRProAlAPheYSeuTRPheII 173
 10 GGAATCCATGCGAGANATGATCTCTGCTTCTGCTGTGTCAT 59
 173 eGLYHIS..... 175
 60 AGGCAATATACATCAATTTATGGATATAGGCAATATACCATCTCC 109
 175 175
 110 TGAGGCTTGTGATTTCTATGTTATGCCGGTGTATATGATGTTAT 159
 176Asn.ArgMetTRPArgLYSAsnArgSerPheT 186
 160 GAACACTACTATGAAAAAGATGATGAGAGAAACCGTTCTTCT 209
 186 YRAlaAsnAsnHISCYsIIeGLYThrasPLeuAsnATGAsnPheAlaSer 202
 210 ATGCAACAATCATTTGATCGATCGAAGACCTGAAATAGAACTTGTCTC 259
 203 LysHISTrpCYsGLuGLuLYAlaSerSerSerSerSerGLuTRTY 219
 260 AACACATGCTGTGAGAGGTGATCCAGTTCCTCATGCTCGAAACCTA 309
 219 RYsGLYLeuTYRProGLuSeRGLuProGLuValLYAlaValAlaSerP 236
 310 CTGTGACCTTATCTCTGATGAGAACCGAAGTGAGCGATGCTAGTT 359
 236 heLeuARgARgAsnIIeAsnGLInIIeLYsAlaTYRLeSerMetHISer 252
 360 TCTTGAGAGAAATATATACACAGATTAAAGCATATACATCAGATCTCA 409
 253 TYRSeRGLNHISIIeValPheProTYRSeRTYRThARgSeRLYsSeRly 269
 410 TACTCCAGCATATAGTGTTCATATTCATATACAGAGTAAGAACAA 459
 269 sAsPHisGLuGLuSeRLeuValAlaSerGLuAlaValAlaArgAlaIleG 286
 460 AGACCATGAGGAACTGCTCTAGTAGCAGTAGAGGAGTTCGTCTATGTG 509
 286 LuLYSThrSeRLYsAsnThrARgTYRThRHISGLYHISGLYSeRGLuThr 302
 510 AGAAAACTAGTAAAAATACAGATATACATGSCCATGGCTCGAANAAC 559
 303 LeuTYRLeuAlaProGLYGLYGLYAsPAsPTRPLeTYRAsPLeuGLYII 319
 560 TTATACCTAGCTCTGAGGAGGAGGAGATGATATATGATTTGGGCAT 609
 319 eLYSTYSeR.Phe..... 323
 610 CAAATATTCGTTTCAATGAACTTCAGATACGGGACATACGAGATTCT 659
 324ThSeRAsnProPovalGLuLYSLeuLeuProLe 335
 660 TGCAGCGGAGCGTTACATCAAAACCACTGTAGAGAAAGCTTTGGCGCT 709

```

seq_name: /SIDSL/gcgdata/geneseq/genesequ-emb1/NA1995.DAT:AAQ90600
seq_documentation_block:
ID   AAQ90600 standard; DNA, 1215 BP.
AC   MAQ90600;
DT   13-MAR-1996 (first entry)
XX
XX Porcine Tyr-His-met procarboxypeptidase B coding sequence.
DE
KW Procarboxypeptidase B; carboxypeptidase B; Pichia; PCPB;
RN human serum albumin; premating factor alpha; mating factor alpha;
RW proCBP; ds.
XX
OS Sus scrofa.
XX
XX Key Location/Organisms
FH CDS 1..1215
FT /*tag= a
FT /product= Porcine procarboxypeptidase B.
FT
PE W09S14096-A1.
PD 26-MAY-1995.
XX
XX 16-NOV-1994; 94WO-USJ1312.
XX
XX 16-NOV-1993; 93US-0153258.
PR
XX (EDIL ) LILLY & CO ELI.
PA
XX Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
PI Sterner JU, Zhang H;
DR WP1: 1995-200386/26.
P-PSDB: AAR75131.
XX
XX DNA encoding porcine carboxypeptidase B - used for transforming
PT host cells, partic. Pichia species; for prodn. of the enzyme
XX
XX Claim 2; Page 20-21; 34pp; English.
XX
XX The porcine carboxypeptidase B coding sequence can be place in a
CC bacterial or pref. Pichia yeast expression vector. The expression
CC vector further comprises the signal peptide of either human serum
CC albumin (designated pG4D3 - NRRL-B-21029); premating factor alpha
CC (designated pF4489 - NRRL-B-21028); mating factor alpha (designated
CC pF4474 - NRRL-B-21032) or the porcine proCBP signal peptide,
CC (designated pIGD27 - NRRL-B-21027). The method can be used for
CC producing large amounts of porcine carboxypeptidase B and when
CC produced in Pichia yeast, the protein does not need solubilisation
CC or folding. The produced enzyme is then used for pref. cleaving
CC basic residues from the carboxy terminus of proteins.
XX
XX Sequence 1215 BP; 358 A; 300 C; 266 G; 291 T; 0 other:

Alignment_scores:
Quality: 646.50 Length: 361
Ratio: 2.694 Gaps: 6
Percent Similarity: 66.482 Percent Identity: 37.950

Alignment_block:
US-09-980-881-3 x AAQ90600 ..

Align seg 1/1 to: AAQ90600 from: 1 to: 1215

1 PhgInSerGIyGlnValLeuAlaAlaLeuProArGTThSerArgClnVa 17

```

28 TTCGAAAGGGGAGGAGGTCCTTCCTGTCATGCTTGAAGATGAAATGACAT 77
17 IGIValLeuGIAsnLeuThrThrThrTyrcIuIeValLeuTrpGln 34
78 CAGCTTACTCCATAGTGTGGCCAGCACCGAGAGATTCACCTTCGGAAC 125
34 roValThrAlaAspLeuIleValLybLybGluValHispheVal 50
128 CAGATTCGTACACCAATATCAAACTCACAGTACAGTTCCTCGGTG 177
51 AsnAlaSerAspValAspAsnValLybAlaHisLeuAsnValSerGlyI 67
178 AAACGACAAAGATATTTTGGCTGTGGAAGACTTTCGGACGAATGAACT 227
67 eProCySerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
228 ACAATATGAGTACTCTATTAACAACCTGGAGATCTGTGCAGAGCTCAGT 277
84 IeSerAsnAspThrValSerProArgAlaSerAlaSerTyrcIuGln 100
278 TTGCACACAGAGTC.....CGTACAACGTGGACACAGTATGAGAG 318
101 TyrHisSerLeuAsnGluIleTyrcTrpIleGluPheIleThrGluAr 117
319 TACACACAGCGGGAACGATCGAGCGGTGGACTTAAGCAAGTCCACAGTGA 368
117 gHisProAspMetLeuThrLybHisIleGlySerSerPheGluYst 134
369 AAATCCAGACCTCATCTCTCGCACACCCCTGGACACTACATTTTACGAA 418
134 yProLeuTyrcValLeuLybValSerGlyLybGluGlnThrAlaLybAsn 150
419 ACAATATATACCTCCCAAGGTT...GGCAAACTGGACAAATPAACCT 465
151 AlaIleTrpIleAspGlyIleHisAlaArgIuTrpIleSerProAl 167
466 GCCATTTCATGACACTGTGTTCATTCGACAGATGATTTCCATGC 515
167 apHeCyLeuTrpPheIle.....173
516 ATTTTGGCAAGTGTGTGTGAGAGAGGCTGTCTACCTAGATGATAGA 565
173173
566 GTCAATGACAGATTCCTCAACAAGCTAAGCTTTATGCTTGCCTGTG 615
174GlyHis.....AsnArgMetTrpAr 180
616 CTCATATGATGATGGTACATCTACACACTCGAACCAAGAAACCGAATGTGAG 665
180 glybAsnArgSerPheTyrcAlaAsnAsnHisCyIleGlyThrAspLeu 197
666 AAAGACCCCGCTCAACAAGCTGGAACTACCTGATTTGGCACACAGCCCA 715
197 snArgAsnPheAlaSerLybHisTrpCySGluGluGluAlaSerSer 213
716 ACAGAAATTTT...CATGTGGGTGTGTGCACAACTGGAGCCTTACAGAC 762
214 SerCySerGluThrTyrcGlyLeuTyrcGluTyrcProGluSerGluProGluVa 230
763 CCGTCGGAGTAGACTTACGTGATCTGCTCACAGACTGCAAAAGAGAC 812
230 IlybAlaValAlaSerPheLeuArgAsnIleAsnGlnIleLybAla 247
813 CAAGGCCCTGGCGATTTATACGCAACAACCTGCTCTCATCAAAAGCAT 862
247 yIleSerMetHisSerTyrcSerGlnHisIleValAPheProTyrcTyrc 263
863 ACCTGACGATCCACATACCTACAGAGATGATACCTTATTCCTAT 912
264 ThrArgSerLybSerLybAspHisGluIleuSerLeuValAlaSerG 280


```

180 glysasnargserphtyrAlaasnanscysIleGlythrAspleua 197
|||||
696 AAGACTCCCTCCACCACCTAGCTGATGGCATGGCAGACCCCA 745
|||||
197 snArgAsnphelaserlyshIstrPCysgluglGlyAlaserSer 213
|||||
746 ACAGAAATTT...CATGCTGTTGGTGAATGGAGCTCTCGAAGC 792
|||||
214 SerCysSerglutThrTyrCysglYleuTyrProgluSergluProgluVa 230
|||||
793 CCCTGATGAAGAACTACTGTGACCTGCCGAGAGTGTGAAGAGAGC 842
|||||
230 lylsAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleYsAlaT 247
|||||
843 CAGAGCCCTGGCTGATTCATCCGACAAACTCTTCATCAAGGCAT 892
|||||
247 yllSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
|||||
893 ATCTGACAAATCCATCTGATCTCCAAATGATGATCTACCTTACTCATAT 942
|||||
264 ThrArgSerLySerLyAsnAspHisgluLeuSerLeuValAlaSerG 280
|||||
943 GCTTACAAACTCGTGAGAACATGCTGATGATGCCCTGGCTAAAGC 992
|||||
280 uAlaValArgAlaIleGluYsThrSerLyAsnThrArgTyrThrHisG 297
|||||
993 TACTGTGAAGAACTT...GCCTCACTGCAGCGCACCAAGTACATATAG 1039
|||||
297 lylHisGlySergluThrLeuTyrLeuAlaProglYglYglYsAspTrp 313
|||||
1040 GCCCGGAGCTCAACAACTATCTCTGCTGCTGGGGGCTCTGACACTGG 1089
|||||
314 lIeTyrAspLeuGlyIleYsTyrSerPheThr 324
|||||
1090 GCTTATGACCAAGAAATGATGATTCCTTCACG 1122
|||||
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV41795
seq_documentation_block:
ID AAV41795 standard; DNA: 1263 BP.
AC
XX
AAV41795;
XX
DT 20-NOV-1998 (first entry)
XX
DE Human pancreatic carboxypeptidase nucleotide sequence.
XX
KW ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
XX
KW produg therapy.
XX
OS Homo sapiens.
FH
FT Key Location/Qualifiers
FT CDS 1..1248
FT /tag- a
FT /product- "Pancreatic carboxypeptidase B"
FT /note- "No start codon given"
XX
MO09835988-A1.
XX
PD 20-AUG-1998.
XX
PF 10-FEB-1998; 98MO-GB00415.
XX
PR 29-OCT-1997; 97GB-0022727.
PR 14-FEB-1997; 97GB-0003104.
PR 18-OCT-1997; 97GB-0022003.
XX
PA (ZENE ) ZENEPRA (ZENE ) ZENECA LTD.
PI Edge MD;
XX
XX
WPI: 1998-467168/40.

```

```

DR P-PSDB; AAW74476.
XX
PT New modified pro-domain of carboxy-peptidase B - enhances expression
PT of co-expressed proteins for production of recombinant
PT carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT enzyme produg therapy
XX
XX
PS Example 1; Page 52-53; 83pp; English.
PS
XX
CC The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC modified pro-domain of CPB on a separate gene to enhance recombinant
CC expression. This process can be used to produce recombinant CPB in
CC eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
CC insulin production and protein sequencing, while its fusions with
CC antibodies are useful in antibody-directed enzyme produg therapy. The
CC modified pro-domain provide increased yields of recombinant CPB, possibly
CC by protecting the C-terminus against enzymatic degradation or by
CC increasing intracellular trafficking.
XX
SQ Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;
|||
alignment_scores:
Quality: 637.50 Length: 361
Ratio: 2.602 Gaps: 6
Percent Similarity: 67.867 Percent Identity: 36.011
alignment_block:
US-09-980-881-3 x AAV41795 ..
Align seg 1/1 to: AAV41795 from: 1 to: 1263
|||
1 PheGlnSerGlyGlnValIleuAlaAlaLeuProArgThrSerArgGlnVa 17
|||||
58 TTGTAAGCGAGAAAGGTCTTCGCTGTAACCTGAAAGTGAATGAAATCAT 107
|||||
17 lGlnValIleuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnP 34
|||||
108 TTAACATATCCCGCGAGTGGCCAGACGACCAAGATTCATTCGGAAGC 157
|||||
34 rovalThrAlaAspLeuIleValYsYsGlnValHisPhePheVal 50
|||||
158 CAGATTCTGTCAACAATCAAACTCACAGTACAGTGAAGTTCGCTGTT 207
|||||
51 AsnAlaSerAspValAlaAspAsnValYsAlaHisIleAsnValSerGly 67
|||||
208 AAAGCAGAAAGTACTGTCATCTGAGAAATGTCTTAAGCAGATGAACT 257
|||||
67 eProCysSerValIleuAlaAspValGluAspLeuIleGlnGlnI 84
|||||
258 ACAATACAGAGTACTGATTAAGCAACCTGAGAAATGTGTGAGGCTCAGT 307
|||||
84 lSerAsnAspThrValSerProArgValAlaSerAlaSerTyrTyrGlnGln 100
|||||
308 TTGATAGCCGGGTT.....CGTGCACAGACACAGTTAAGAGAG 348
|||||
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluAr 117
|||||
349 TACAACAAGTGGAAAGCATGAGAGCTTGAGCTCAACAAGTGGCAGCTGA 398
|||||
117 gHisProAspMetLeuThrIlyslleHisIleGlySerSerPheGluYsT 134
|||||
399 GAATCCAGCCCTCATCTCTGCACTGTTATCGGAAACCATTTGAGAGAC 448
|||||
134 YRPLeuTyrValIleuYsValSerGlyYsGluGlnIleThrAlaYsAsn 150
|||||
449 GCGGTATTTACTCTTAAGGTT...GCCAAAGCTGACAAAATTAAGCCT 495
|||||
151 AlaIleTrpIleAspCysglYleHisAlaArgGluTrpIleSerProAl 167
|||||
496 GCCATTTTCATGAGACTGTGTTCCATGCCAGAGAGATTCCTCTCTGTC 545
|||||
167 apheCysLeuThrPheIle..... 173

```



```

|||||
546 ATTCCTGCCAGTGGTTGTAGAGAGCGCTGTTCTACCTATGAGCTGAGA 595
FT
173 ..... 173
596 TCCAACTGACAGAGCTTCGACAGTTAGACTTTATGTCGCCCTG 645.
174 .....GlyHis.....AsnArgMetTrpPar 180
646 CTCATATATGCTAGCTACCTACCTGACCAAGACCGATTTTGAG 695
180 glyAsnArgSerPheThrAlaAsnAsnHisCysIleGlyThrAspLeu 197
696 AAAGACTCGCTCCACCATCTGATCTACCTGATTCGATGGACAGACCCCA 745
197 snArgAsnPheAlaSerLyHisIleTrpCysGluGluValAlaSerSer 213
746 ACAGAAATTTT...GATGCTGGTTGGTGTGAATGTGAGCCTTCGAAC 792
214 SerCysSerGluThrTyrcysglyleuTyrrProGluSerGluProGlu 230
793 CCTGTGATGAACCTACTGTGACCTGCCGACAGCTCTGAAAAGAAAC 842
230 IlyAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAla 247
843 CAAGGCCCTGGTGAATTCATCCGCAACAACTCTTCATCATCAAGCAT 892
247 ylleSerMetHisSerTyrrSerGlnHisIleValPheProTyrrSerTy 263
893 ATCTGACAACTCCTCTACTCCCAATGATGATCTACCTTACTCATAT 942
264 ThrArgSerLySerLyAspHisGluLeuSerLeuValAlaSerG1 280
943 GCTTACAAATCGCTGAGAACATGCTGATGAATCCCTGGCTAAAGC 992
280 uAlaValAlaGlnIleGluLyThrSerLyAsnThrArgTyrrThrIsg 297
993 TACTGTGAAGAACT...CCCTACTGACGACGACCAACTACATATG 1039
297 IlyHisGlySerGluThrLeuTyrrLeuAlaProGlyGlyAspAspTrp 313
1040 GCCCGGAGCTACACAACTATCTCTGCTGCTGGGGCTGTGACGACTGG 1089
314 IleTyrrAspLeuGlyIleLysTyrrSerPheThr 324
1090 GCTTATGACCAAGAAATCAGATATTCCTCACC 1122
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AA142506
seq_documentation_block:
ID AA142506 standard; DNA; 1284 BP.
AC AA142506;
XX
XX
DT 17-FEB-1997 (first entry)
XX
XX
DE ProHCPB gene with Pe1b leader sequence.
XX
XX
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPPT;
KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Fd; F(ab')2; Pe1b; leader; human carboxypeptidase B; ss.
XX
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1275
FT /tag= a
FT sig_peptide 1..66
FT /tag= b
FT /note= "Pe1b leader sequence"
FT sig_peptide 67..351

```

```

FT
FT mat_peptide /tag= c
FT /note= "pro sequence"
FT 352..1272
FT /tag= d
FT /note= "mature HCPB sequence"
XX
XX W09620011-A1.
XX
XX 04-JUL-1996.
XX
XX 21-DEC-1995; 95WO-GB02991.
XX
XX 16-AUG-1995; 95GB-0016810.
XX 23-DEC-1994; 94GB-0026192.
XX
XX (ZENE ) ZENECA LTD.
XX
XX Blakey DC, Boyle PT, Davies DH, Eggelte HJ, Heaton DW,
XX Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM,
XX Tarragona-Fiol A, Taylorson CJ;
XX WPI: 1996-321650/32.
XX P-PSDB: AAM06175.
XX
XX Two component system for anti-tumour therapy - comprising targeting
XX moiety linked to mutated enzyme which can transform an
XX anti-neoplastic prodrug
XX
XX Reference Example 18; Page 140-142; 182pp; English.
XX
XX A two-component system for anti-tumour therapy comprises a targeting
XX moiety linked to a mutated enzyme which can transform an anti-neoplastic
XX prodrug. The system is based on antibody directed enzyme prodrug therapy
XX (ADPPT) using non-naturally occurring mutant forms of host enzymes,
XX pref. human pancreatic ribonuclease (HP-RNase). (see AA142478-83).
XX Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
XX be used. The present sequence is a cloned pro-HCPB gene contained in
XX plasmid pIC11738 and which can be expressed in E. coli.
XX
XX
XX Sequence 1284 BP; 355 A; 315 C; 299 G; 315 T; 0 other;
SQ
alignment_scores:
Quality: 637.50 Length: 361
Ratio: 2.602 Gaps: 6
Percent Similarity: 67.867 Percent Identity: 36.011
alignment_block:
US-09-980-881-3 x AA142506 ..
Align seg 1/1 to: AA142506 from: 1 to: 1284
1 PheGlnSerGlyGlnValIleuAlaLeuProArgThrSerArgGlnVal 17
|||||
85 TTGTAAAGCCGAGAGGTGTTCGTTTAAAGCTTAAAGCAATGCAAT 134
17 IglValIleuGlnAsnLeuThrThrThrTyrrGluIleValLeuTrpGln 34
|||||
135 TAACATATATCCGAGATTGGCCACAGACCACTGATGACTTCGGAAGC 184
34 roValThrAlaAspLeuIleValIlyslYsgIlnValIHisPheVal 50
|||||
185 CAGATTCTGTCCACCAACAACTCAACGCTAAGCTGATGATCCGCTGT 234
51 AsnAlaSerAspValAlaAspAsnValIleHisIleuAsnValSerGlyI 67
|||||
235 AAACGAGAGACTACTGTCACTGTGAGAGATGTTTAAAGCAGATGAACT 284
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
|||||
285 ACAATACAGAGTACTGATTAACCAACCTGAGAAATGTGTGAGGCTCAGT 334
84 IeSerAsnAspThrValSerProArgAlaSerAlaSerTyrrGluGln 100

```

```

335 TTTGATACCCGGGTT.....CGTGCAACAGGACACACTTATGAGAG 375
101 TTTTATSerLeuAnGluIleuTyrSerTrpIleGluPheIleThrGluAr 117
376 TACACACAGTGGAGAACGATAGAGCGGTGGACATCAACAATGCCACTGA 425
117 gHISProAspMetLeuThrIleuHISIleGlySerSerPheGluLysTr 134
426 GAATCCAGCCCTCATCTCGCAGTGTATATCGAGACCACTTTGAGGAGAC 475
134 YrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
476 GCGCTATTATACCTCTCGAAGCTT...GCCAAAGCTGACAAATAAGCCCT 522
151 AlaIleTrpIleAspCysGlyIleHISAlaArgGluTrpIleSerProAl 167
523 GCCATTTTCATGGACATGCTGGTTCCATGCCAGAGCTGATTTCTCTCTGC 572
167 aPheCysLeuTrpPheIle..... 173
573 ATTCTGCAGTGGTTGTAGAGAGCGCTTTCGTACCTATGAGAGCTGAGA 622
173 ..... 173
623 TCCCAAGTGAAGAGCTTTCGACAACTTATGACTTTATGCTCGCTGTG 672
174 .....GlyHIS.....AsnArgMetTrpAr 180
673 CTCCAATATTGATGGCTATACATCTACACCTGACCAAGAGCCGATTTGGAG 722
180 gHISAsnArgSerPheTyrAlaAsnAsnHIScysIleGlyTrpAspLeuA 197
723 AAGAGCTCGCTCCACCATCTAGCATGTAGCTGATGGCAGAGAGCCCA 772
197 snArgAsnPheAlaSerLysHISTrpCysGluGluGluLysAspSerSer 213
773 ACAGAAATTTT...GATCGTGGTGGTGTGAATGGACACCTCTCGAAGAC 819
214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVa 230
820 CCGCTGAGAGAAACTTACTGACCTGCCGACAGTCTGAAAGAGAGAC 869
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaAr 247
870 CAAAGCCCGCGCTGATTTCATCCGACAAACATCTCTTCATCAAGGCAAT 919
247 YrIleSerMetHISerTyrSerGlnHISIleValPheProTyrSerTyr 263
920 ACTGACAAATCCACTGCTACTCCAAATGATGATGATACCCCTTACATCAT 969
264 ThrArgSerLysSerLysAspHISGluGluLeuSerLeuValAlaSerGl 280
970 GCTTTCAAACATCGGTGAGAACATCTGATGTGAATGCCCTGGCTAAAGC 1019
280 uAlaValAlaArgAlaIleGluLysThrSerLysAsnThrArgTyrTrpHISG 297
1020 TACTGTGAAGAAAGCTT...GCCCTACTGACGCGACCAAGTACATATNG 1066
297 LysHISGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAspTrp 313
1067 GCCCGGAGGCTTACAAACATCTATCTCTGCGCGGCTCTGACAGCTGG 1116
314 IleTyrAspLeuGlyIleLysTyrSerPheTr 324
1117 GCTTATGACCAAGGAAATCATGATTTCTTCAAC 1149

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAD03837

seq_documentation_block:

ID AAD03837 standard: cDNA: 1311 BP.

XX AAD03837:

```

XX XX
DT 19-JUN-2001 (first entry)
XX XX
DE Human carboxypeptidase-related protein #5 CDNA.
XX XX
KM Human; carboxypeptidase-related protein; protease; gene therapy;
XX XX drug screening; mental; biological disorder; medical disorder; ss.
OS Homo sapiens.
XX XX
FH Key Location/Qualifiers
FT CDS 1..1311
FT FT /*tag= a
FT FT /product= "Human carboxypeptidase-related protein #5"
FT FT /note= "CDS does not include stop codon"
XX XX /partial
PN WO200123588-A1.
XX XX
PD 05-APR-2001.
XX XX
PF 29-SEP-2000; 2000WO-US26876.
XX XX
PR 29-SEP-1999; 99US-0156685.
XX XX
PA (LEXI-) LEXICON GENETICS INC.
XX XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX XX WP1: 2001-266171/27.
XX XX P-PSDB; AAE00502.
XX XX
PT New isolated human carboxypeptidase polynucleotide useful for
PT generating antibodies, as reagents in diagnostic assays and for
PT screening for compounds useful for treating mental, biological or
PT medical diseases
XX XX
PS Claim 1; Page 30; 36pp; English.
XX XX
CC The present sequence is a cDNA encoding novel human protein (NHP),
CC carboxypeptidase-related protein. The carboxypeptidase-related
CC proteins share structural similarity with animal proteases, and
CC especially carboxypeptidase B or carboxypeptidase A. The
CC carboxypeptidase-related protein cDNA is useful for the detection of
CC mutant human carboxypeptidases, or inappropriately expressed human
CC carboxypeptidases for the diagnosis of disease, and also as a
CC therapeutic. It is also useful in drug screening, for generation of
CC antibodies, as reagents in diagnostic assays, for the identification of
CC reagents in assays for screening for compounds that are useful for
CC treating mental, biological or medical disorders and diseases. Nucleotide
CC constructs encoding functional NHPs can also be used in gene therapy.
XX XX
SQ Sequence 1311 BP; 391 A; 278 C; 284 G; 358 T; 0 other;

```

alignment_scores:
 Quality: 619.50 Length: 391
 Ratio: 2.560 Gaps: 9
 Percent Similarity: 61.893 Percent Identity: 34.783

alignment_block:
 US-09-980-881-3 x AAD03837 ..

Align seg 1/1 to: AAD03837 from: 1 to: 1311

```

1 PheGlnSerGlyGlnValLeuAlaLeuProArgTrpSerArgGlnVa 17
: : : : : : : : : : : : : : : : : : : : : : : : : :
109 TATGCTGCTGATAAGTATAGATTATTCCTCAAAACAGAGAGGAGAC 158
17 LglnValLeuGlnAsnLeuThrTrpTrpTyrGluIleValLeuTrpGln 34
: : : : : : : : : : : : : : : : : : : : : : : : : :
159 ATATGCACTGAAGAAATATCTTATCAACTTAAAGCTGAGACTGTGGAGC 208

```


alignment_block:
US-09-980-881-3 x AAO90601 ..

Align seg 1/1 to: AAO90601 from: 1 to: 921

```

98 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIle 114
|||||.....:|||||.....:|||||.....:
16 TATGAGAAAGTACAAACCTGGGAACGATGAGGCTTGGACTTAAGCAAC 65
|||||.....:|||||.....:|||||.....:
114 ethrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerP 131
|||||.....:|||||.....:|||||.....:
66 CACCAAGTGAATAATCCAGACCTCATCTCTGCACAGCAATCGGAATCACT 115
|||||.....:|||||.....:|||||.....:
131 heGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThr 147
|||||.....:|||||.....:|||||.....:
116 TTTTAGCAAAACAATATATACCTCTCAAGTTT...GCCAAACCTGGACCA 162
|||||.....:|||||.....:|||||.....:
148 AAlaLysAsnAlaIleTrpIleAspGlyIleHisAlaArgIleTrpIle 164
|||||.....:|||||.....:|||||.....:
163 AATAAGCCCTGCCATTTTCATGGAGCTGGTTCCATGCCAGAGAAATGAT 212
|||||.....:|||||.....:|||||.....:
164 eSerProAlaPheCysLeuTrpPheIle.....:|||||.....: 173
|||||.....:|||||.....:|||||.....:
213 TTCCCATGATTTTGGCAGTGTGTGTGAGAGAGGCTGTCTCAGCTATG 262
|||||.....:|||||.....:|||||.....:
173 .....:|||||.....:|||||.....: 173
|||||.....:|||||.....:|||||.....:
263 GATATGAGAGTCACATGACAGAAATTCCTCAACAGCTAGACTTTTATGTC 312
|||||.....:|||||.....:|||||.....:
174 .....:|||||.....:|||||.....: 174
|||||.....:|||||.....:|||||.....:
313 TTGCTGTGCTCAATATGATGGCTACATCTACACCTGGACCAAGACCG 362
|||||.....:|||||.....:|||||.....:
177 gMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyT 194
|||||.....:|||||.....:|||||.....:
363 AATGTGGAGAAAGACCCGCTCTACCAATGCTGGAGACTACCTGCATTTGCA 412
|||||.....:|||||.....:|||||.....:
194 hArgPheLysAsnArgAsnPheAlaSerLysHisIleTrpCysGluGluLysAla 210
|||||.....:|||||.....:|||||.....:
413 CAGACCCCAACAGAAATTTT...GATGCTGGGTGGTGCACAAACTGAGACC 459
|||||.....:|||||.....:|||||.....:
211 SerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerG 227
|||||.....:|||||.....:|||||.....:
460 TCTACAGACCCCTGGATGAGACTTACGTGATGTGCTGCAGAGTCTGCA 509
|||||.....:|||||.....:|||||.....:
227 uProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnI 244
|||||.....:|||||.....:|||||.....:
510 AAAAGAGCAAGGCCCTGGCTGATTTTATAGCAACCACTCTCTCCA 559
|||||.....:|||||.....:|||||.....:
244 lAlaValAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
|||||.....:|||||.....:|||||.....:
560 TCMAAGCATACCTGACGATCCATCTATCTACAGATGATACTACCT 609
|||||.....:|||||.....:|||||.....:
261 TyrSerTyrTrpArgSerLysSerLysAspHisGluGluLeuSerLeuVal 277
|||||.....:|||||.....:|||||.....:
610 TATTCCTGTGATTAACAACCTCCCGAGAAACAATGCTGAGTTGAATTAACCT 659
|||||.....:|||||.....:|||||.....:
277 lAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgT 294
|||||.....:|||||.....:|||||.....:
660 GGCTTAAGGCTGCCGTGAAGAACTT...GCTACACTGTATGAGCAACAGT 706
|||||.....:|||||.....:|||||.....:
294 YrTrpHisGluHisIleGlySerGluThrLeuTyrLysAlaProGlyGly 310
|||||.....:|||||.....:|||||.....:
707 ACAATATACGGCCCAAGAGACTTACAACAATCTATCTCTGCTGGGGGCTCT 756
|||||.....:|||||.....:|||||.....:
311 AspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThr 324
|||||.....:|||||.....:|||||.....:
757 GATGACTGGGCTTATGACCAAGAAATCAAAATATCTCTTCAACC 798
|||||.....:|||||.....:|||||.....:

```

seq_name: /STDSJL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA03839

seq_documentation_block:
ID AAD03839 standard; DNA: 2128 BP.

```

XX XX AAD03839;
AC XX
DE XX 19-JUN-2001 (first entry)
DE XX
DE XX Human carboxypeptidase-related protein DNA with flanking 5' and 3' ends.
DE XX
DE XX Human; carboxypeptidase-related protein; protease; gene therapy;
KW drug screening; mental; biological disorder; medical disorder; ds.
XX
OS Homo sapiens.
XX
XX WO200123588-A1.
XX
XX 05-APR-2001.
XX
XX 29-SEP-2000; 2000WO-US26876.
XX
XX 29-SEP-1999; 99US-0156685.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
XX Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-266171/27.
XX
XX New isolated human carboxypeptidase polynucleotide useful for
XX generating antibodies, as reagents in diagnostic assays and for
XX screening for compounds useful for treating mental, biological or
XX medical diseases -
XX
XX Disclosure; Page 33; 36pp; English.
XX
XX The present sequence is a novel human protein (NHP),
XX carboxypeptidase-related protein DNA with 5' and 3' flanking ends. The
XX carboxypeptidase-related proteins share structural similarity with
XX animal proteases, and especially carboxypeptidase B or
XX carboxypeptidase A. The carboxypeptidase-related protein cDNA is useful
XX for the detection of mutant human carboxypeptidases, or inappropriately
XX expressed human carboxypeptidases for the diagnosis of disease, and also
XX as a therapeutic. It is also useful in drug screening, for generation of
XX antibodies, as reagents in diagnostic assays, for the identification of
XX other cellular gene products related to human carboxypeptidases, and as
XX reagents in assays for screening for compounds that are useful for
XX treating mental, biological or medical disorders and diseases. Nucleotide
XX constructs encoding functional NHPs can also be used in gene therapy.
XX
XX Sequence 2128 BP; 640 A; 461 C; 436 G; 591 T; 0 other;

```

alignment_scores:
Quality: 571.50 Length: 441
Ratio: 2.342 Gaps: 13
Percent Similarity: 55.329 Percent Identity: 31.746

alignment_block:
US-09-980-881-3 x AAD03839 ..

Align seg 1/1 to: AAD03839 from: 1 to: 2128

```

3 SerGlyGlnValLeuAlaIleLeuProArgThrSerArgGlnValGlnVal 19
|||||.....:|||||.....:|||||.....:
461 AGTGAATAAGTGAATAGATTATTTCCAAAACAGAGAGCAATGATGC 510
|||||.....:|||||.....:|||||.....:
19 lLeuGlnAsnLeuThr-ThrThyGluIleVal.....: 30
|||||.....:|||||.....:|||||.....:
511 ACTAAGAAATATATCTATCACTTAAGTTGGTTCTGCACATCAAGTGC 560
|||||.....:|||||.....:|||||.....:
31 ...LeuTrpGlnProValThrAlaAspLeuIleValLys.....LysLys 44
|||||.....:|||||.....:|||||.....:
561 GACCTGTGGCAGCCCAAGAGATCTCTATGTATCAAGAGGACAGACTTAC 610
|||||.....:|||||.....:|||||.....:
44 sGlnValHisPhePheValAsnAlaSerAspValAlaAspAsnValLysAla 61
|||||.....:|||||.....:|||||.....:

```


This Page Blank (uspto)

OM of: US-09-980-881-3 to: Issued_Patents_NA:* out_format : pfs

Date: Sep 18, 2002 6:55 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frim+...p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US09980881.runat.16092002.140129.10653/app.query.fasta.1.1126
-DB=Issued_Patents_NA -OPT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPT=0.000 -LOOEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -FGAPOP=10.000 -FGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=10sum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORR=ext -HEARISE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 -CGN1_1.62 -NCPU=6 -ICPU=3 -LONGLOC
-DEV=TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-3
Query length: 338
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 90.520000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	..	1272	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-869-057-1 +		1732.00	3887.62	7.5e-209	1272	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-07-649-591B-2 +		1732.00	3883.43	1.3e-208	1749	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-277-540-2 +		1732.00	3883.43	1.3e-208	1749	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-430-787A-2 +		1732.00	3883.43	1.3e-208	1749	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-696-139-1 +		646.50	1426.09	9.6e-72	1215	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-56 +		637.50	1405.17	1.4e-70	1263	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-71 +		637.50	1404.95	1.4e-70	1284	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-171-945-124 +		619.50	1357.31	6.5e-68	2154	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-782-760-5 +		610.00	1346.87	2.5e-67	927	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-869-057-1 +		610.00	1346.87	2.5e-67	927	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-696-139-3 +		582.00	1283.44	8.4e-64	921	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-67 +		567.00	1248.35	7.6e-62	999	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-64 +		567.00	1247.66	8.3e-62	1053	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-74 +		560.00	1231.70	6.4e-61	1059	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-77 +		559.00	1229.43	8.6e-61	1059	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-171-945-112 +		549.00	1199.26	4.1e-59	1870	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-395-936-3 +		516.50	1130.84	2.7e-55	1251	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-395-936-1 +		491.50	1074.07	3.9e-52	1257	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-395-936-1 +		491.50	1074.07	3.9e-52	1257	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-331-709-2 +		210.50	446.65	3.4e-17	591	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-996-416-125 +		169.50	355.47	4.1e-12	515	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-484-105-15 +		116.50	214.42	0.0003	2504	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-484-105-15 +		116.50	214.42	0.0003	2504	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-706-216-1 +		102.00	180.44	0.0231	2719	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-782-760-2 +		97.00	198.81	0.0022	285	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-869-057-1 +		97.00	198.81	0.0022	285	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-07-915-203-1 +		94.00	157.90	0.4164	3797	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-277-540-2 +		94.00	157.90	0.4164	3797	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-789-449-1 +		94.00	157.90	0.4164	3797	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-609-049A-29 +		91.00	146.74	1.74	5285	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-170-996-29 +		91.00	146.74	1.74	5285	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-609-049A-27 +		91.00	143.36	2.69	6831	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-170-996-27 +		91.00	143.36	2.69	6831	1		
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-073-237-1 +		90.50	163.75	0.1367	1333	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-689-974-2 -		88.50	157.01	0.4669	1576	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-09-058-376-2 -		88.50	157.01	0.4669	1576	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-09-233-989-1 +		88.50	151.26	0.9765	2439	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-452-262-1 +		88.50	151.23	0.9792	2443	1		
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-734-550-1 +		88.50	151.23	0.9792	2443	1		

/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-869-057-1 + 88.50 151.23 0.9792 24
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-412-102-5 + 87.50 155.41 0.5733 1498
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-217-787-5 + 87.50 155.41 0.5733 1498
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-254-358-1 - 86.00 137.00 6.08 4680
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-475-391-1 - 86.00 137.00 6.08 4680

seq_name: /cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-869-057-1

seq_documentation_block:

Sequence 1, Application US/08869057
Patent No. 5985562
GENERAL INFORMATION:
APPLICANT: Morser, Michael J
TITLE OF INVENTION: Method of Detecting Thrombotic Disease
TITLE OF INVENTION: Risk
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Herlex Biosciences Legal Department
STREET: 15049 San Pablo Avenue
CITY: Richmond
STATE: California
COUNTRY: USA
ZIP: 94804-0099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,057
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Washien, Wendy L.
REGISTRATION NUMBER: 36,301
REFERENCE/DOCKET NUMBER: 51509AUSMI
TELEPHONE: 510-262-5411
TELEFAX: 510-262-7095
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1272 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PUBLICATION INFORMATION:
AUTHORS: Eaton, Dan L.
AUTHORS: Malloy, Beth E.
AUTHORS: Tsai, Siac P
AUTHORS: Henzel, William
AUTHORS: Drayna, Dennis
TITLE: Isolation, Molecular Cloning, and Partial
TITLE: Characterization of a No. 5985562 Carboxypeptidase B
JOURNAL: J. Biol. Chem.
VOLUME: 266
ISSUE: 32
PAGES: 21833-21838
DATE: No. 5985562 15-1991

US-08-869-057-1

alignment_scores:

Quality	1732.00	Length:
Ratio	5.124	Gaps: 3
Percent Similarity	86.005	Percent Identity: 85.751

alignment_block:

US-09-980-881-3 x US-08-869-057-1 ..
Align seg 1/1 to: US-08-869-057-1 from: 1 to: 1272

```

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgLys 17
   |||
67 TTTTCAGAGTGGCCAGATCTTACCTGCTTCTCTAGAACCTTAGGCAAGT 116
   |||
17 LglnValLeuGlnAsnLeuThrThrThrArgLysValLeuLeuProGln 34
   |||
117 TCAGATTCTACGAATCTTACTACCAACATATGAGATTGTTCTCTGCGACG 166
   |||
34 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
   |||
167 CGGTACACAGTACCTTATTGTGACAAAAAACATCCATTTTCTTGTGA 216
   |||
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGly 67
   |||
217 AATGCATCTGATGTCGACAAATGGAAGCCCATTTAAATGTGAGCGCAAT 266
   |||
67 eProCysSerValLeuLeuAlaAspValGlnAspLeuIleGlnGlnThr 84
   |||
267 TCCATGCACTGCTTGTCTGCGACGCTGGAGATCTTATTCAACACGACA 316
   |||
84 LeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGln 100
   |||
317 TTTCCAAAGCACACAGTCAAGCCCCGAGCCTCCGCACTGACTATGACAG 366
   |||
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 117
   |||
367 TATCACTCACTAAATGAAATCTATCTTGATAGAAATTTATTAACGACAG 416
   |||
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
   |||
417 GCATCCTGATGCTTACAAAAATCCACATCTGATCCTATTGAGAACT 466
   |||
134 YrProLeuTyrValLeuLysValSerGlyLysGlnGlnThrAlaLysAsn 150
   |||
467 ACCCACTCTATGTTTAAAGTTCTGGAAAGAACAAACACCAAAAT 516
   |||
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
   |||
517 GCCATATGATGACGTGGAATCCATGCGACAGAAATGATCTCTCCGTC 566
   |||
167 aPheCysLeuTrpPheIleGlyHis..... 175
   |||
567 TTTTCGCTTGTGTTCATAGGCCATATACTCAATTCTATGGGATAAAG 616
   |||
175 ..... 175
   |||
617 GGCAATATACCAATCTCCTGAGGCTTGTGATTTCTATGTCGCGGTG 666
   |||
176 .....AsnArgMetTrpArg 180
   |||
667 GTTAAATGCGACGTTATGACTACTCATGAGAAAAAAGATCGAATGTGAG 716
   |||
180 gLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeu 197
   |||
717 AAAGAACCGTCTTCTTATGAGAACATCATGATCGAAGACAGACGTA 766
   |||
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSer 213
   |||
767 ATAGGAACCTTGTCTTCAAAACACTGCTGAGGAGGTCATCCAGTCC 816
   |||
214 SerCysSerGluThrTyrCysGlyLeuTyrProGlnSerGluProGlnVal 230
   |||
817 TCATCTCTGGAAACCTTACTGTGACTTATCTCTGCTGAGAACCGAAGT 866
   |||
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAla 247
   |||
867 GAAGGCACTGCGCTAGTCTTGAGAAAGAAATATCAACAGATTAAAGCAT 916
   |||
247 YrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
   |||
917 ACATCAGCATCATTCATATCTCCAGCATATAGTTCATATATCTCAT 966

```

```

264 ThrArgSerLysSerLysAspHisGlnGluLeuSerLeuValAlaSerG 280
   |||
967 ACACGAAGTAAAGCAACACCATGAGGAAGTCTCTAGTACCCAGTGA 1016
   |||
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrHisG 297
   |||
1017 AGCAGTTCGTGCTATGAGAAAACTAGTAAAAATACCAAGTATACAGT 1066
   |||
297 LysIleGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAspTrp 313
   |||
1067 GCCATGCTTCAGAAACCTTATACCTAGCTCTGAGAGTGGGACGATTGG 1116
   |||
314 IleTyrAspLeuGlyIleLysTyrSer.Phe..... 323
   |||
1117 ACTATGATTGGGCAATCAATATTTGTTTACAATTGAACCTCGAGATAC 1166
   |||
324 .....ThrSerAsnProGluVal 329
   |||
1167 GGCAATACATGAGATCTTGTCTCCGAGCGCTTACATCAAAACCCACTGA 1216
   |||
330 GlnLysLeuLeuProLeuSerLeuLys 338
   |||
1217 GAGAACTTTTCCGCTGCTCTTAAAA 1243

```

seq_name: /cgn2_6/plodata/1/ina/5A_COMB.seq:US-07-649-591B-2

seq_documentation_block:

Sequence 2, Application US/07649591B

Patent No. 5206161

GENERAL INFORMATION:

APPLICANT: Dennis Drayna and Daniel Eaton

TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/649,591B

FILING DATE: 19910201

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Hasak, Janet E.

REGISTRATION NUMBER: 28,616

REFERENCE/DOCKET NUMBER: 689

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/266-1896

TELEFAX: 415/952-9881

TELEX: 910/371-7168

SEQUENCE CHARACTERISTICS:

INFORMATION FOR SEQ. ID NO. 2:

LENGTH: 1749 bases

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: hybridization probe

LOCATION: 133 to 178

IDENTIFICATION METHOD:

OTHER INFORMATION:

FEATURE:

NAME/KEY: potential clip site

```

; LOCATION: 380 to 382
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: signal sequence
; LOCATION: 41 to 106
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-07-649-591B-2

```

```

alignment_scores:
    Quality: 1732.00    Length: 393
    Ratio: 5.124        Gaps: 3
    Percent Similarity: 86.005    Percent Identity: 85.751

```

```

alignment_block:
US-09-980-881-3 x US-07-649-591B-2

```

```

Align seg 1/1 to: US-07-649-591B-2 from: 1 to: 1749

```

```

1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnVal 17
107 TTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTGAACTCTAGCAAGT 156
17 LglnValLeuGlnAsnLeuThrThrThrGluIleValLeuTrpGlnP 34
157 TCAAGTTCTACAGAACTTCTACTACACATATGAGATTGTCTCTGCGAGC 206
34 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
207 CGGTAACAGGTGACCTTATTTGTGAAGAAAAACAAGTCCATTTTGTGA 256
51 AsnAlaSerAspValAlaAspValAlaHisLeuAsnValSerGlyI 67
257 AATGCATCTGATGCGACAAATGTCAAGCCCATTTAATGCGAGCGAAT 306
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
307 TCCATGCGAGTGTCTTGTGCGAGCGTGAAGATCTTATTCACACAGAG 356
84 leSerAsnAspThrValSerProArgAlaSerAlaSerTyrrGlyGln 100
357 TTTTCAACGACAGACAGTACGCCCGCGAGCTCTCGCATGCTACTATGAACAG 406
101 TyrHisSerLeuAsnGluIleTyrrSerTrpIleGlnPheIleThrGlu 117
407 TATCACCACCAATGAATCTATCTTGGATAGCAATTTAATGAAGAGAG 456
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
457 GCATCCCGAATATGCTTCAAAAAATCCACATTTGATCTCATTTGAGAGT 506
134 yrrProLeuTyrrValLeuLysValSerGlyLysGlnIleThrAlaLys 150
507 ACCCAGCTGATGTTTAAAGTTTCTTGAAAAAGAACCAACGCAAAAAAT 556
151 AlaIleTrpIleAspCysGlyIleHisAlaArgLutTrpIleSerProAl 167
557 GCCATATGATGATGACTGTGGAATCCATGCCAGAGAAATGGAATCTCTCC 606
167 apheCysLeuTrpPheIleGlyHis..... 175
607 TTTCTGCTTGTTGTTCAATAGCCCATTAATCAATCTATGATGATTAATAG 656
175 ..... 175
657 GGCATATTAACCAATCTCTGAGGCTTGATTTCTATGTTATGCCGCTG 706
176 .....AsnArgMetTrpPar 180
707 GTTAATGTGACGGTTATGACTACTCATGAAAAAGAAATCAATGATGTGAG 756

```

```

180 glysAsnArgSerPheTyrrAlaAsnAsnHisCysIleGlyThrAspLeuA 197
757 AAAGAACGGTCTTCTATGATGGAACAATCATTTGATGGAACAGACCTGA 806
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGluAlaSerSerSer 213
807 ATAGAACTTGTCTTCCAAACACTGTGTGTGAGAAAGGTGCTACCAAGTCC 856
214 SerCysSerGluThrTyrrCysGlyLeuTyrrProGluSerGluProGlu 230
857 TCATGCTCGGAAACCTACTGCTGAGCTTATCTGATGAGAACAGAAAGT 906
230 llysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAla 247
907 GAAGCAGGTGGCTAGTTCTTGAAGAAAGAAATATCAACCGATTAAGCAT 956
247 yrrIleSerMetHisSerTyrrSerGlnHisIleValPheProTyrrSerTy 263
957 ACATCAGCATGATCATCTACTCCAGCATATATGATGTTCCATATTCCTAT 1006
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
1007 ACAGAAAGTAAAGCAAAAGACCATGAGGAAGTGTCTAGTAGCAGTGA 1056
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrrHisG 297
1057 AGCAGTTGCTGCTATTTGAGAAACTAGTAAATAACAGATATACATG 1106
297 lYHisGlySerGluThrLeuTyrrLeuAlaProGlyGlyGlyAspAspTr 313
1107 GCCATGCTCAGAAACCTTATACCTAGCTCCTGAGAGGTGGGAGCATG 1156
314 lIleTyrrAspLeuGlyIleLysTyrrSer.Phe..... 323
1157 ATCTATGATTTGGGCATCAATATATCTGTTACATTTGACATTCAGATAC 1206
324 .....ThrSerAsnProPheVal 329
1207 GGGCACAATACGAAATCTTGTCTGCGGAGCGTTACATCAAAACCACTGA 1256
330 gLutLysLeuLeuProLeuSerLeuLys 338
1257 GAGAAAGCTTTGCCGCTGCTCTTAAAA 1283
seq_name: /cgn2_6/prodata/1/lna/5A_COMB.seq:us-08-277-540-2
seq_documentation_block:
; Sequence 2, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET INFORMATION:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-277-540-2

```

```

Alignment scores:
Quality: 1732.00      Length: 393
Ratio: 5.124          Gaps: 3
Percent Similarity: 86.005      Percent Identity: 85.751

```

```

Alignment block:
US-09-980-881-3 x US-08-277-540-2

```

```

Align seg 1/1 to: US-08-277-540-2 from: 1 to: 1749

```

```

1 PheGInSerGlyGlnValLeuAlaLeuProAlGthSerArgGlnVa 17
107 TTTCAGAGGCGCAAGTCTAGCTCTCTCTCTAGAACCTTAGCGCAGT 156
157 TCAGATTCTACAGAACTTACTACACAAATAGATGTCTCTGCGCAGC 206
17 IGInValLeuGlnAsnLeuThrThrThyGlnIleValLeuTyrPglInp 34
157 TCAGATTCTACAGAACTTACTACACAAATAGATGTCTCTGCGCAGC 206
34 rovalThAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
207 CGGTACAGCGCTGACCTTATTGTGAAGAAAAACAAGTCATTTTGTGA 256
51 AsnAlaSerAspValAlaAspValLysAlaHisLeuAsnValSerGlyTl 67
257 AATGATCTGATGTGACAAATGTGAAGCCCATTTAAATGTGAGCGGAGT 306
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
307 TCATGACAGTGTCTGTGCGAGAGCTGGAAGATCTTATTCACAGCAGA 356
84 IeSerAsnAspThrValSerProAlGAlaSerAlaSerTyrGlnGln 100
357 TTTCACAGCAGACAGTCAGCCCGGAGCTCCGCACTCTACTATCAACAG 406
101 TyrHisSerLeuAsnGlnIleTyrSerTrpIleGluPheIleThGluAr 117
407 TATCACTCTACAAATGAATCTATCTTGATAGATTAATTAAGTGAAG 456
117 GHISProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysT 134
457 GCATCTGATATGCTTACAAAAATCCACATGATGATCCATTTGAGAGT 506
134 yTrProLeuTyrValLeuLysValSerGlyLysGlnGlnThAlaLysAsn 150
507 ACCCACTCTATGTTTAAAGGTTTCTGAAAAAGAACAAAGCCAAAAAT 556
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
557 GCCATATGAGATTGACTGTGGAATCCATGCCAGAAATGATCTCTCTGC 606
167 apheCysLeuTrpPheIleGlyHis..... 175

```

```

|||||
607 TTTCCTCTGTGCTTCATAGCCATATACCAATCTATGGGATTAATAG 656
175 ..... 175
657 GGCAATATACCAATCTCTAGAGCGTTTGATTTCTATGTATGCCGGTG 706
176 .....AsnArgMetTrpAr 180
707 GTTAATGTGAGGGTTATGACTACTACATCGAAAAAGAAATGCAATGTGAG 756
180 gLysAsnArGserPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuA 197
757 AAAGAACCGTCTTCTATGCGAACAAATCATATGATCGAACAGACGTGA 806
197 snArgAsnPheAlaSerLysHisIleTyrCysGlnGlnGluAlaSerSer 213
807 ATAGGACTTTCCTTCCAAACAGTGTGTGAGGAAGGTCCATCCAGTTCC 856
214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVa 230
857 TCATGCTCGAAGAACTACTGTGACTTATCCTGATCGAAGACGAAAGT 906
230 lLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlat 247
907 GAAGGACGTGCTAGTCTTCTGAGAAAGAAATATCAACCAATTTAAAGCAT 956
247 yTrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
957 ACATGACATGATCATCATCTCCAGCATATATGTTTCCATATTTCCAT 1006
264 ThrArgSerLysSerLysAspHisGlnGluLeuSerLeuValAlaSerG 280
1007 ACACGAAGTAAAGCAAGAACATGAGAACGTCCTCTAGTAGCCAGTGA 1056
280 uAlaValArgAlaIleGlyLysThrSerLysAsnThrArgTyrThrHisG 297
1057 AGCACTTGTCTATGTGAGAAACTATGTAATAAATACAGGTATACATG 1106
297 lYHisGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAspTrp 313
1107 GCCATGGCTCAGAAACCTTATACCTAGCTCCTGAGAGTGGGAGCATTTGG 1156
314 lLeTyrAspLeuGlyIleLysTyrSer.Phe..... 323
1157 ATCATGATTTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGATAC 1206
324 .....ThrSerAsnProProVal 329
1207 GGGCAGATACGGAATTTGCTGCCGAGACGTTACATCAAAACCCACGTGA 1256
330 GluLysLeuLeuProLeuSerLeuLys 338
1257 GAGAAAGCTTTGCCGCTGTCTTAAAA 1283

seq_name: /cgn2_6/prodata/1/lna/5A_COMB.seq:US-08-430-787A-2

seq_documentation_block:
; Sequence 2, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674del Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genetech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEX: 415/952-9881
FAX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-430-787A-2

alignment_scores:
Quality: 1732.00      Length: 393
Ratio: 5.124          Gaps: 3
Percent Similarity: 86.005      Percent Identity: 85.751

alignment_block:
US-09-980-881-3 x US-08-430-787A-2
Align seg 1/1 to: US-08-430-787A-2 from: 1 to: 1749

1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnA 17
|||||
107 TTTCAGAGTGGCCAGATTCTAGCGCTCTCTAGAACCTCTAGCGCAGT 156
17 GlnValLeuGlnAsnLeuThrThrThrTyrgLulleValLeuTrpGlnP 34
|||||
157 TCAGATTCTACAGAAATCTTACTACACATATGAGATGTTCTCTGCGCAGC 206
34 roValThrAlaAspLeuIleValIleLysLysGlnValAlaHisPhePheVal 50
|||||
207 CGGTAAACGCTGACCTTATTGTGAGAAAAAACAGAGTCCATTTTGTGTA 256
51 AsnAlaSerAspValAspAsnValIleAlaHisLeuAsnValSerGlyI 67
|||||
257 AATGCATCTGATGTCGACAAATGTGAAGCCCATTTAATGTGACCGCAAT 306
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnI 84
|||||
307 TCCATGCGAGTGTCTGTCTGCGACGCGTGAAGATCTTATTCACACAGAGA 356
84 leSerAsnAspThrValSerProArgAlaSerAlaSerIleTyrgLugln 100
|||||
357 TTTCACACGACACAGTACGCCCGCGAGCTCCCGCATGCTACTATGACACAG 406
101 TyrHisSerLeuAsnGluIleTyrgSerTrpIleGluPheIleThrGluAr 117
|||||
407 TATCACTACACTAAATGAATCTATTTTGGATGAAATTTTAATTAAGTAGAG 456

```

```

117 gHisProAspMetLeuThrIleHisIleGlySerSerPheGluLysT 134
|||||
457 GCATCGCTGATATGCTTACAAAAATCCACATGATGATCTCATTTAGAGAT 506
134 YrProLeuTyValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
|||||
507 ACCCATCTATGTTTAAAGGTTTCTGCAAAAGAACAAACAGCCAAAT 556
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
|||||
557 GCCATATGATGATGACTGTGGAATCCATGCGAGAAATGAGATCTCTGCG 606
167 aPheCysLeuThrPheIleGlyHis..... 175
607 TTTCGCTGTGTTGATAGCCATATATCAATCTATGAGTAATAG 656
175 ..... 175
657 GGCATATACCAATCTCCTGAGGCTTGATGATTTCTATGTTATCCGCGTG 706
176 .....AsnArgMetTrpPar 180
707 GTTAATGTGACGGTATATGACTACTCATGAAAAAAGAAATGAAATGTGGAG 756
180 GlyAsnArgSerPheTyrgAlaAsnAsnHisCysIleGlyThrAspLeuA 197
|||||
757 AAAGAACCGTCTTCTATGCGAAATCATTCGATGCGAAGACCTGA 806
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGluValAlaSerSer 213
807 ATAGGAATTTGCTTCCAAACACTGTGTGAGAGAGGCGATCCAGTTC 856
214 SerCysSerGluThrTyrgCysGlyLeuTyrgProGluSerGluProGluA 230
|||||
857 TCATGCTCGGAAACCTACTGTGACTTATCTAGTCAGACAGACAGAT 906
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlat 247
907 GAAGCAGGTGCTGATTCTTGAGAAAGAAATATACCCAGATTTAAACAT 956
247 YrIleSerMetHisSerTyrgSerGlnHisIleValPheProTyrgSerTy 263
957 ACATCAGCATGATTCATTACTCCACAGATATAGTTCATATTCCTAT 1006
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
1007 ACAGAAAGTAAAGCAAAAGACCATGAGGAATGCTCTAGTAGCAGATGA 1056
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrgThrHisG 297
1057 AGCAGTTCGTCTATTGAGAAACTAGTAAAAATACAGATATACATG 1106
297 LysIleGlySerGluThrLeuTyrgLeuAlaProGlyGlyLysAspTrp 313
1107 GCCATGGCTCGAAGACCTTATTACTTAAGCTCTGAGGAGGAGGAGCATTTG 1156
314 IleTyrgAspLeuGlyIleLysTyrgSer.Phe..... 323
1157 ATCTATGATTTGGGCATCAATATATTCGTTTACATTTGACTTGAGATAC 1206
324 .....ThrSerAsnProProVal 329
1207 GGGCACAATACGAAATCTGTGTCGCGGAGCGTTACATCAATAAACCCAGTGA 1256
330 GluLysLeuLeuProLeuSerLeuLys 338
1257 GAGAAAGCTTTTGCCGCTGCTCTTAATAA 1283
seq_name: /cgn2_6/plodata/1/lna/5A.COMB.seq:US-08-696-139-1
seq_documentation_block:
; Sequence 1, Application US/08696139
; Patent No. 5672496

```

178 AAGCACAAGATATTTTGGCTGGGAAGACTTTCTGGAGCAAGATGAACACT 227
67 eProcySerValIleuLeuAlaSerValGluAspLeuIleGlnGlnI 84
228 ACAATATGAGGTATCTTAAACAACCCGAAATCTGTGCTGGAGCTCACT 277
84 IeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
278 TTGACACAGAGTC.....CGTCAACTGGACACAGTTATGGGAAG 318
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluAr 117
319 TACAACAACCTGGGAACGATCGAGCGCTTGCACTACCAACAAGTCCACGTGA 368
117 gHisProAspMetLeuThrTyrHisIleHisIleGlySerSerPheGluLysTr 134
369 AAATCCGAGACTATCTCTCTGGCAGACGACCTGGGAACCTACATTTTATGGA 418
134 yProLeuTyrValIleuLysValSerGlyLysGluGlnThrAlaTyrAsn 150
419 ACAATATATACCTCCCTCAAGGT...GGCAACCTGGACCAATATAGCT 465
151 AlaIleThrPheAspProGlyIleHisAlaArgGluTrpIleSerProAl 167
466 GCCATTTTCAATGAGCAGTGGTTTCATGCGACAGAAATGATTTCCCATGC 515
167 aPheCysIleuTrpPheIle..... 173
516 ATTTTGGCAGTGGTTTGTGAGAGAGCGCTGTCTCAACTATGGATATAGAGA 565
173 173
566 GTCACATGACAGAAATCTCTCAACACAGCTAGTTATGCTTGGCTGTG 615
174GlyHis.....AsnArgMetTrpAr 180
616 CTCGAATATTGATGGCTACACTCTACACCTGGACACAGAACCGAATGGAG 665
180 GlyAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAr 197
666 AAAGACCCGCTCTACCAAGCTGGAACCTACCTGCATTGGCACAGACCCCA 715
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGluValAspSerSer 213
716 ACAGAAATTTT...GATGCTGGGTGGTGACAACTGGAGACCTCTCAACAG 762
214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVal 230
763 CCGTCGATGAGACTTACTGTGGATCTGGTCGACAGAGTCTGAAAAAGAC 812
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTr 247
813 CAAGGCGCTGGCTGATTATATCGGACAAACCTCTCTCCATCAACAAAGCAT 862
247 yTrIleSerMetHisSerTyrSerGlnHisIleValAlaProTyrSerTyr 263
863 ACCTGACGATCCACTCATCTGACAGATGATCTACTACCTTATTCCTAT 912
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280
913 GATTCACAACTCCCGAGACAACTAGCTAGTTGAATTAACCTGGCGTAAGGC 962
280 ValAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisG 297
963 TGCCTGGAAAGAACTT...GCTACACTGTATGGCCACCAAGTCACTATACG 1009
297 LysHisGlySerGluThrLeuTyrLeuAlaProGluGlyGluLysAspArgTr 313
1010 GCGCAGAGAGCTCAACAACTCTATCTGCTGGGAGGCTGTGATGATACGCG 1059
314 IleTyrAspLeuGlyIleLysTyrSerPheThr 324
1060 GCTTATGACCAAGGAATCAAAATATCTCTTCAACC 1092


```

seq_name: /cgn2_6/ptodata/1/ina/58_COMB.seq:US-08-860-882A-71

seq_documentation_block:
  Sequence 71, Application US/08860882A
  Patent No. 5985281
  GENERAL INFORMATION:
    APPLICANT: TAYLORSON, CHRISTOPHER JOHN
    APPLICANT: EGGELOE, HENDRIKUS JOHANNES
    APPLICANT: TARRAGONA-FLOL, ANTONIO
    APPLICANT: RABIN, BRIAN ROBERT
    APPLICANT: BOYLE, FRANCIS THOMAS
    APPLICANT: HENNAM, JOHN FREDERICK
    APPLICANT: BLAKELY, DAVID CHARLES
    APPLICANT: MARSHAM, PETER ROBERT
    APPLICANT: HEATON, DAVID WILLIAM
    APPLICANT: DAVIES, DAVID HOW
  TITLE OF INVENTION: CHEMICAL COMPOUNDS
  NUMBER OF SEQUENCES: 77
  CORRESPONDENCE ADDRESSES:
    ADDRESSEE: PILLSBURY, MADISON & SUTRO
    STREET: 1100 NEW YORK AVENUE, N.Y.
    CITY: WASHINGTON
    STATE: D.C.
    COUNTRY: USA
    ZIP: 20005
  COMPUTER READABLE FORM:
    MEDIUM TYPE: FLOPPY Disk
    COMPUTER: IBM compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/860,882A
    FILING DATE: JUNE 23, 1997
    CLASSIFICATION: 424
    ATTORNEY/AGENT INFORMATION:
      NAME: DONALD J. BIRD
      REGISTRATION NUMBER: 25,323
      REFERENCE/DOCKET NUMBER: 9901/238653
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: (202) 861-3027
      TELEFAX: (202) 822-0944
      TELEX: 6174627 CUSH
    INFORMATION FOR SEQ ID NO: 71:
      SEQUENCE CHARACTERISTICS:
        LENGTH: 1284 bases
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
  US-08-860-882A-71

Alignment_scores:
  Quality: 637.50      Length: 361
  Ratio: 2.602         Gaps: 6
  Percent Similarity: 67.867      Percent Identity: 36.011

Alignment_block:
  US-09-980-881-3 x US-08-860-882A-71  ..

Align seg 1/1 to: US-08-860-882A-71 from: 1 to: 1284

1 PhcGInserGyGInValLeuAlAlaLeuProArGThrSerArGInVa 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 TTTGAAGCGGAAAGCTTCGCTGATTACGTGAGATGAAATACAT 134
17 GInValLeuGInAsnLeuThrTrhTrhTyGluIlleValLeuTrpGln 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 TACATATATCCGCGAGGTGGCCAGCAGACCAACCAATTCGACTTCGGAAGC 184

```

34 rovalThralaaspleuileValuyluyluysolnValnhsPheReval 50
185 CAAATTCTGTGCACAAATCAATCAACCTGCACGATGATTCCTCCGTGT 234
51 AsnAlaSerApValaLaspValValysAlaValnhsLsAsnValSerGlyL 67
235 AAGCGAAGAAATGATCTGCACTGTGGAGAAATGTTCTTAAGACAAATGACT 284
67 eProCysSerValLeuLeuAlaLaspValGluAspleuileGlnGlnGln 84
285 ACAATACAAAGATCTATTAAGCAACCTGGAAATGTGGTGAGGCGCACT 334
84 leSerAsnAspThrValSerProArgAlaSerAlaSerAlaSerTyrGluGln 100
335 TTGATGACCGCGGT.....CGTGCACAGACAGACAGTATGAGAG 375
101 TyrHisSerLeuAsnGluileTyrSerThrPrilleGluPheileThGluAr 117
376 TACAAACAATGGGAAACGATAGAGGTGGATGCATCAACAATGGCCACATGA 425
117 gHisProAspMetLeuThrLysLileHisileGlySerPheGluTyr 134
426 GAATCAGGCGCCATCTCTCCGACGTTATTCGGAACCATTTGAGGAC 475
134 yProLeuTyrValLeuLysValSerGlyLysGluGlnThralaLysAsn 150
476 GCGCTATTTCACCTCCGAAAGGT...GGCAAAAGCTGGACAAATTAAGCT 522
151 AlaileThrilleAspCysGlyileHisAlaArgGluThrLileSerProAl 167
523 GCCATTTTATGAGCATGCTGGTTCCATGCCAAGAGATGGATTTCTCTGC 572
167 abPecysleuTyrPheile..... 173
573 AATCTGCCAGTGATGTTGTAAGAGAGGCTGTCTGCTATGAGACGTGAG 622
173 173
623 TCCAACTGACAGAGCTTCTGCACAAAGTTAGACTTTTATGTCTCGCTGTG 672
174 GlyHis.....AsnArgMetThrPar 180
673 CTCATATTGATGAGGCTACATCTACAACCTGGACAGAGCCGATTTGGAG 722
180 gLysAsnArgSerPheTyrAlaAsnAsnHisCysLileGlyThrAspleuA 197
723 AAGAGCTGCTCTCCACCCATACGATCGATTCAGCTGGACAGAGCCCA 772
197 snArgAsnPheAlaSerLysHisTyrCysGluGluGlyAlaSerSerSer 213
773 ACAGAAATTTT...GATGCTGGTTGTGTGAATTTGAGACCTCTCGAAGAC 819
214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVa 230
820 CCGCTGTGATGAACCTACTGTGGACCTCGCGAGAGTGTGAAGAGAGAC 869
230 LysAlaValaLaserPheLeuArgArgAsnLleAsnGlnLileLysAlAT 247
870 CAAAGCCCTGGTGATTTCAATCCGCAACAACCTCTTTCATCAAGGGAT 919
247 yLileSerMetHisSerTyrSerGlnHisileValaPheProTyrSerTyr 263
920 AATCTGACAAATCCACGTCGATCCCAATAGATGATCTACCTTACATCAT 969
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGcl 280
970 GCTTACAAACTCGGAGAACAAATGCTAGTTGAATGCCCGGCTAAAGC 1019
280 uAlaValArgAlaileGlyLysLysSerLysAsnThrArgTyrThrHisG 297
1020 TACTGTGAAGAAGACT...GGCTACCTGCACGGCACAAGTACACATATG 1066
297 LysHisGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAspAspTr 313


```

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0336/43847-A-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 927 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..927
PCT-US96-00995-5

```

```

alignment_scores:
Quality: 610.00 Length: 269
Ratio: 3.389 Gaps: 5
Percent Similarity: 66.914 Percent Identity: 44.981

```

```

alignment_block:
US-09-980-881-3 x PCT-US96-00995-5

```

```

Align seg 1/1 to: PCT-US96-00995-5 from: 1 to: 927

```

```

93 Alaseralaserlytyrlygluyltyrlyhlseryleuansgluletyrse 109
1 GCAAGTGGACACAGCTACACCAAGTACCAACTGGAAACGATTGAGGC 50
109 rtrpilleglupheilerthgluathlsproaspmetleuthlystle 126
51 GTGATTCACAAAGTTGCCACTGATATATCCAGACTGTCACACAGCG 100
126 lslllelyserpnehlulstyrproleuthlyvalleulysvalser 142
101 TCATTTGGACACATTGGAAGCGTACATGTATGCTCAAGATT... 147
143 glylysgluylthrlyalysasnalaltrpilespercysglylehl 159
148 GGTAAACCTAGACCGAATAGCGTCGCATCTCATCGATTGTGTTCCA 197
159 salaatglutrpilesrproalaphcysleutrpheile... 173
198 TGCAGAGAGTGGATTTCCTCATTCTGTCAGTGTGTGAGAGAG 247
173 ..... 173
248 CTGTCCGCTACTTAAATCAAGAGATCCACATGAAACGCTTACATGAA 297
174 .....glyhls..... 175
298 CTGGATTTCATGTCTGCCTGTGGTCAACATTCATGATGCTTCACAC 347
176 .....AsnatrmettrparlygsasnatrgserphtyrAlaAsna 189
348 CTGGACTAGAGACAGATGTGGAGAAAACCCGCTACTACTATGCTGGA 397
189 snhlsCyslllelythraspleuansnargasnphelaaserlyshstirp 205
398 GTTCCTGCTTGGGTGTAGACCCCAACGGAAATTT...AATGCTGGCTGG 444
206 CysglugluylalaserSerSerSerCysSerCysgluthrtyrCysglyle 222
445 TGTGAAGTGGAGCTTCTCGGAGTCCCTGCTCGAAGCTTACTGTGAGAC 494

```

```

222 utyrprogluserluprogluvallysalavalalaserpheleuarg 239
495 AGCCCGAGAGCTGTAAGAGACAAAGCCCTGCGAATTCATCCGCA 544
239 rgaenileasnglillelysalatryllesermethtlsertyrsergl 255
545 ACACCTCTCCACCATCAAGGCTTACCTGACCATTCACCTACTCTCAG 594
256 hislevalpneprotlyserlytyrthraspserlyserlyshlsigl 272
595 ATGATGCTACCTCTACTCTGATGACTACAAACTGCTGAGAACTATGA 644
272 ugluSerleuValalaserCysgluAlaArgAlalleglulsthrs 289
645 GGAATTAATGCCCTGGTGAAGGTGGCGCAAGACCTT...GCCACTC 691
289 eulyasnthrarglytrthrslyhlslysglysergluthrleuthleu 305
692 TGCATGGACCAAGTACACATATGCCCCAGAGCTTACACATATCTCT 741
306 Alaprogllyglyaspasprpilletyraspleuaglylelystyrse 322
742 GCTGCTGGGGATCTGACGACTGCTTATGATCAGGGAATCAATATTC 791
322 rpherthr 324
792 CTTTACC 798

```

```

seq_name: /cgn2_6/plodata/1/lna/5A_COMB.seq:US-08-696-139-3

```

```

seq_documentation_block:
; Sequence 3, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hersberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Steiner, Jane L.
; APPLICANT: Zhang, Hailiao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 base pairs
; TYPE: nucleic acid

```


67. GCACTGGTCTCTTACGAGAAATACACACAGTGGGAAACATAGAGGC 116
109 rtrp1leGluphe1lethrg1uargh1sProaSPmetleuthr1ys1leH 126
117 TTGGACTCAACAAGTGGCCACATGAGATTCAGCCCTCATCTCTCGCAGTG 166
126 1s1leGlySerSerPheG1uLysTyPProLeuTyPValLeuLysValSer 142
167 TTATCGAAGACACATTTGAGGAGCGCGCTATTACCTCGTGAAGGTT... 213
143 Gly1ySGluGlnThrAlaLysAsnAla1leTrp1leAspCysGly1leH 159
214 GGCAAAAGCTGGACAAAATTAAGCTGCAATTTCAATGAGCTGTGGTTCCA 263
159 sAlaArg1uTrp1leSerProAlaPheCysLeuTrpPhe1le..... 173
264 TGGCAGAGATGGATTCTCTCGATCTGCAAGTGTGTGTAAGAGAG 313
173 173
314 CTGTGTGACCTATGAGAGTGAATCCAGTACAGAGCTTCTGACAG 363
174 GlyHs..... 175
364 TTAGACTTTATGTCTGCTGCTGCTCAATATTGATGGCTATCTACAC 413
176 AsnArgMetTrpArgLysAsnArgSerPheTyra1aAsnA 189
414 CTGGACCAAGAGCGAATTTGGAGAAAGACTCGCTCCACCATCTGAT 463
189 snH1cys1leGlyThrAsPLeuAsnArgAsnPheAlaSerLysH1sTrp 205
464 CTAGCTGCATTGGCACAGACCCCAAGAAATTTT... GATGCTGCTGG 510
206 CysG1uG1uLys1leSerSerSerCysSerG1uThrTyP1CysGlyLe 222
511 TGTGAATTTGGAGCTCTCGAAACCCCTGTGATGAACTTACTGTGGACC 560
222 uTyPProG1uSerG1uProG1uValLysAlaValAlaSerPheLeuArg 239
561 TGGCGGAGACTGTGAAGAAGAGACCAAGGCCCTGCGTATTCATCCGCA 610
239 rGAsn1leAsnG1u1leLysAlaTyP1leSerMetH1sSerTyP1SerG1n 255
611 ACAAGCTCTCTCCATCAAGCATATCTGACAATCCACTGCTCTCCCAA 660
256 H1s1leValPheProTyP1SerTyP1ThrArgSerLysSerLysAspH1sG1 272
661 ATGATGTCTTACCTTACTCATATGCTTACAAACTGGTGAGAACATATGC 710
272 uG1uLeuSer1eUValAlaSerG1uAlaValArgAla1leG1uLysThrs 289
711 TGAATTTGAATGCGCTGCTAAGCTACTGTGAAGAAGCTT... GCGTCAC 757
289 e1rLysAsnThrArgTyP1ThH1sG1yH1sG1ySerG1uThrLeuTyP1leu 305
758 TGCACGCGACCAAGTACATATGCGCGGAGCTACACAACTATATCT 807
306 AlAProG1yLys1yAspAspTrp1leTyP1AsPLeuG1y1leLysTyP1 322
808 GCTGCTGGGGGCTCTGAGACTGGGCTTATGACCAAGAAATCAATATTC 857
322 rPheThr 324
858 CTTCAC 864

seq_name: /cgn2.6/plodata/1/ina/5B.COMB.seq:US-08-860-882A-64
seq_documentation_block:
: Sequence 64, Application US/08860882A
: Patent No. 5985281
: GENERAL INFORMATION:
: APPLICANT: TAYLORSON, CHRISTOPHER JOHN

APPLICANT: EGGELTE, HENDRIKUS JOHANNES
APPLICANT: TARRAGONA-FIOL, ANTONIO
APPLICANT: RABIN, BRIAN ROBERT
APPLICANT: BOYLE, FRANCIS THOMAS
APPLICANT: HENNAM, JOHN FREDERICK
APPLICANT: BLAKELY, DAVID CHARLES
APPLICANT: MARSHAM, PETER ROBERT
APPLICANT: DAVIES, DAVID HUM
TITLE OF INVENTION: CHEMICAL COMPOUNDS
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILSBURY, MADISON & SUTRO
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 1053 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-860-882A-64

alignment_scores:
Quality: 567.00 Length: 269
Ratio: 3.133 Gaps: 5
Percent Similarity: 67.286 Percent Identity: 41.264

alignment_block:
US-09-980-881-3 x US-08-860-882A-64 ..
Align seg 1/1 to: US-08-860-882A-64 from: 1 to: 1053

93 AlAserAlaSerTyTyG1uG1uThH1sSerLeuAsnG1u1leTyP1Ser 109
117 TTGGACTCAACAAGTGGCCACTGAGAAATCCAGCCCTCATCTCTCGCAGTG 166
109 rtrp1leGluphe1lethrg1uargh1sProaSPmetleuthr1ys1leH 126
117 TTGGACTCAACAAGTGGCCACTGAGAAATCCAGCCCTCATCTCTCGCAGTG 166
126 1s1leGlySerSerPheG1uLysTyPProLeuTyPValLeuLysValSer 142
167 TTATCGAAGACACATTTGAGGAGCGCGCTATTACCTCGTGAAGGTT... 213
143 Gly1ySGluGlnThrAlaLysAsnAla1leTrp1leAspCysGly1leH 159
214 GGCAAAAGCTGGACAAAATTAAGCTGCAATTTCAATGAGCTGTGGTTCCA 263
159 sAlaArg1uTrp1leSerProAlaPheCysLeuTrpPhe1le..... 173
264 TGGCAGAGATGGATTCTCTCGATCTGCAAGTGTGTGTAAGAGAG 313

```

173 ..... 173
314 CTGTTGTAACCTATGAGACGTGAGATCCAAAGTACAGAGCTTCTGACAAAG 363
174 ..... GlyHis..... 175
364 TTAGACTTTTATGTCCTGCTGCTGCTCAATATTGATGCTTACTTACAC 413
176 ..... AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsn 189
414 CTGACCAAGAGCCGATTTGGAGAAAGACTGCTCCACCACTACAGAT 463
189 snHisCysIleGlyThrAspLeuAsnArgAspPheAlaSerLysHisTrp 205
464 CTAGCTGCAATGGCAGACACCCCAAGAAATTTT...GATGCTGCTGG 510
206 CysGluGluGlyAlaSerSerSerSerSerGluThrTyrCysGlyLe 222
511 TGTGAATGAGAGCCCTCTGAAACCCCTGTGATGAATCTTACTGTGACC 560
222 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 239
561 TCCCGCAGAGCTGAAAGAGAGACCAAGCCCTGGCTGATTCATCCGCA 610
239 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 255
611 ACAAACTCTCTTCATCAAGAGCATATGACAAATCCACATCCGACCCCA 660
256 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGln 272
661 ATGATGATCTACCTTACTCATATGCTTACAAACATCGGAGAACAAATGC 710
272 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysHis 289
711 TGAGTTGATGCCCTGCTAAAGCTAGCTGTAAGAAGACTT...GCCCTCAC 757
289 eTyrAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 305
758 TGCACGGACCAAGTACACATATGACCGGAGCTACAAACAATCATCTCT 807
306 AlaProGlyGlyLysAspArgTrpIleTyrAspLeuGlyIleLysTyrSe 322
808 GCTGCTGGGGGCTGACGAGCTGGCTTATGACCAAGATCAGATATTC 857
322 rPheThr 324
858 CTTCACC 864
seq_name: /cgn2_6/prodata/1/lna/5h_COMB.seq:us-08-860-882A-74
seq_documentation_block:
Sequence 74, Application US/08860882A
Patent No. 5985281
GENERAL INFORMATION:
APPLICANT: TAYLORSON, CHRISTOPHER JOHN
APPLICANT: EGGELTE, HENDRIKUS JOHANNES
APPLICANT: TARRAGONA-FIOL, ANTONIO
APPLICANT: RABIN, BRIAN ROBERT
APPLICANT: BOYLE, FRANCIS THOMAS
APPLICANT: HENNAM, JOHN FREDERICK
APPLICANT: BLAKELY, DAVID CHARLES
APPLICANT: MARSHAM, PETER ROBERT
APPLICANT: HEATON, DAVID WILLIAM
APPLICANT: DAVIES, DAVID HIR
TITLE OF INVENTION: CHEMICAL COMPOUNDS
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA

```

```

ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1059 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-860-882A-74

Alignment_scores:
Quality: 560.00 Length: 269
Ratio: 3.111 Gaps: 5
Percent Similarity: 66.914 Percent Identity: 40.892

Alignment_block:
US-09-980-881-3 x US-08-860-882A-74 ..
Align seg 1/1 to: US-08-860-882A-74 from: 1 to: 1059

93 AlaSerAlaSerTyrTyrGluGlnIleTyrHisSerLeuAsnGluIleTyrSe 109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 GCACGTGCTACTCTTACGACAAATCAACAAAGTGAAGCATGAGAGCC 116
109 TrpIleGluPheIleThrGluArgHisProAspMetLeuThrIleHis 126
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 TTGCACTCAACAAGTGGCCACTGAGAAATCCAGCCCTCATCTCTGCACTG 166
126 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TTATCGAACAACATTTGAGGAGCGCGCTATTACTCTCTGAGGTT... 213
143 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 GCAAAAGCTGACAAATAATAGCTGCCATTTTCATGACCTGGTTCACA 263
159 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle..... 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TGCACAGAGAGGATTTCTCTGCAATTTGCCAGTGGTTGTAAAGAGAGG 313
173 ..... 173
314 CTGTTGTAACCTATGAGACGTGAGATCCAAAGTACAGAGCTTCTGACAAAG 363
174 ..... GlyHis..... 175
364 TTAGACTTTTATGTCCTGCTGCTGCTCAATATTGATGCTTACTTACAC 413
176 ..... AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsn 189
414 CTGACCAAGAGCCGATTTGGAGAAAGACTGCTCCACCACTACAGAT 463
189 snHisCysIleGlyThrAspLeuAsnArgAspPheAlaSerLysHisTrp 205
464 CTAGCTGCAATGGCAGACACCCCAAGAAATTTT...GATGCTGCTGG 510

```

```

206 CysGlulGluGlyAlaSerSerSerSerSerGluThrTyrcysgyle 222
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
511 TGGAATTGGAGCTCTCGAAACCCCTGATGAACCTTACGTGGACC 560
222 uTYrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 239
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
561 TGCCGAGAGTCTGAAAAGAGACCAAGCCCTGGCTGATTCATCCGCA 610
239 rGAsnIleAsnGlnIleLysAlaIleTyrIleSerMetHisSerTyrSerGln 255
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
611 ACAACTCTCTCCATCAAGCATATCTGACAATCCACTGCTACCCAA 660
256 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 272
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 ATGATGATCTACCTTACTATATGCTTACCAACTCGGTGAGAACAAATGC 710
272 uGluLeuSerLeuValAlaSerGluValAlaArgAlaIleGluLysThr 289
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
711 TGAGTTGAATGCCCTGGCTAACTACTGTGAAGAAGACTT..GCCCTCAC 757
289 eRLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 305
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
758 TGCACGACCAAGTACATATATGCGCCGGAGCTACACAACTATACCT 807
306 AlaProGlyGlyLysAspTrrPleTyrAspLeuGlyLysTyrSe 322
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
808 GCGTCGGGGGCTTAAGACTGGGCTTATGACCACCAAGAAATCAGATATTC 857
322 rPheThr 324
      |||||
858 CTTCACC 864

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-860-882A-77

```

seq documentation block:
; Sequence 77, Application US/08860882A
; Patent No. 5983281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGELE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUM
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944

```

```

; TELER: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1059 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-860-882A-77

alignment_scores:
      Quality: 559.00      Length: 269
      Ratio: 3.106      Gaps: 5
      Percent Similarity: 66.914      Percent Identity: 40.892

alignment_block:
US-09-980-881-3 x US-08-860-882A-77
Align seg 1/1 to: US-08-860-882A-77 from: 1 to: 1059

93 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGlnIleTyrSe 109
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 GCAACTGTCTACTCTTACGAGAAATACACAAAGTGGAGGAACGATAGAGCC 116
109 rTrrPleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 126
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 TTGACCTCAACAAGTCCGACCTGAGAAATCCAGCCCTCATCTCTGCACTG 166
126 IsIleGlySerSerPheGluLysTyrProLeuTyrValIleLysValSer 142
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 TTATCGAACCACATTTGAGGAGCGCTATTTACTCTCGAAGTT... 213
143 GlyLysGluGlnThrAlaLysAsnAlaIleTrrPleAspCysGlyIleH 159
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 GGCNAAGCTGCAAAATTAAGCCCTGCCATTTTCATGAGACTGTGTTCCA 263
159 sAlaArgGluTrrPleSerProAlaPheCysLeuTrrPheIle..... 173
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
264 TGCAGAGAGAGGATTTCTCTGCTATTCGCCAGTGGTTGTAGAGAGG 313
173 ..... 173
314 CTGTTCCTACCTATGAGCTGAGATCCAACTGACAGACTTCTGCAGAG 363
174 .....GlyHis..... 175
364 TTACACTTTTATGTCGCGCTGCTCAATATGATGCTACACTAC 413
176 .....AsnArgMetTrrPArgLysAsnArgSerPheTyrAlaAsnA 189
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
414 CTGCACCAAGAGCCGATTTTGGAAAGACCTGCCACCATCTGAT 463
189 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrr 205
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
464 CTACCTGATTTGGCAGACAGCCCAAGAAATTTT...GATGCTGGTGG 510
206 CysGluGluGlyAlaSerSerSerSerSerSerGluThrTyrcysgyle 222
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
511 TGGAATTGGAGCTCTCGAAACCCCTGATGAACCTTACGTGGACC 560
222 uTYrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 239
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
561 TGCCGAGAGTCTGAAAAGAGACCAAGCCCTGGCTGATTCATCCGCA 610
239 rGAsnIleAsnGlnIleLysAlaIleTyrIleSerMetHisSerTyrSerGln 255
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
611 ACAACTCTCTCCATCAAGCATATCTGACAATCCACTGCTACCCAA 660
256 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 272
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 ATGATGATCTACCTTACTATATGCTTACCAACTCGGTGAGAACAAATGC 710
272 uGluLeuSerLeuValAlaSerGluValAlaArgAlaIleGluLysThr 289

```


OM of: US-09-980-881-3 to: EST: * out_format : pfs
 Date: Sep 18, 2002 5:41 PM
 About: Results were produced by the GenCore software, version 4.5.
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=framed-p2n.model -DEV=x1h
 -Q=/cgr1_1/USPRO.spool/US09980881/runat.16092002.140128.10623/app_query.fasta.1.1126
 -DB=EST -QMT=fastap -SUFFIX=st -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPEXT=4.500
 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blomsom2 -TRANS=human40.cdl
 -LIST=45 -DOCLLOCAL=200 -THR_SCORE=pt -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=6 -MAXLEN=200000000 -USER=US09980881 @CGNL 1.3434
 -NCPU=6 -ICPU=3 -IONCLOC -DEV=TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPRY -WAIT -THREADS=1

Search information block:
 Query: US-09-980-881-3
 Database: EST: *
 Database sequence: 13736207
 Database length: -1841457050
 Search time (sec): 3115.820000

Score_list:	Strd Orig	ZScore	EScore	Len	Documentation
Sequence					
gb_hhc:AK004045	+ 1494.00	2980.16	9.5e-157	1446	AK004045 Mus musculus 18 days
gb_est1:BL144664	+ 1244.50	2483.01	4.7e-129	889	BL144664 AL536069 IFLI.FL013.FBI
gb_est2:BG569281	+ 988.00	1966.63	2.7e-100	853	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 978.00	1946.49	3.6e-99	852	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 926.50	1844.45	1.7e-93	788	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 924.00	1835.58	5.4e-93	1022	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 921.00	1832.52	8.0e-93	788	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 912.00	1815.01	7.5e-92	747	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 897.00	1784.78	6.6e-90	747	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 894.50	1780.10	6.6e-90	724	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 863.00	1716.46	2.3e-86	735	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 862.00	1714.60	2.9e-86	725	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 831.50	1653.74	7.2e-83	688	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 828.50	1645.88	6.0e-82	806	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 814.50	1618.50	6.6e-81	750	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 812.50	1617.52	7.5e-81	575	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 806.50	1603.83	4.4e-80	661	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 806.00	1604.32	4.1e-80	580	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 806.00	1604.15	4.2e-80	589	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 785.00	1559.99	1.2e-77	691	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 776.50	1544.43	8.9e-77	603	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 766.00	1525.18	1.0e-75	511	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 759.00	1511.07	6.4e-75	511	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 755.00	1501.50	2.2e-74	583	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 751.50	1490.20	9.3e-74	843	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 744.00	1478.96	3.9e-73	602	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 741.00	1472.94	8.5e-73	601	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 735.00	1460.43	4.2e-72	623	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 732.50	1457.10	6.5e-72	537	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 727.50	1443.10	3.9e-71	749	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 725.00	1440.32	5.6e-71	621	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 715.00	1416.19	1.2e-69	958	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 707.00	1404.26	6.6e-68	723	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 698.00	1385.17	5.5e-67	487	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 688.00	1368.56	2.1e-66	772	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 676.50	1341.35	1.8e-65	692	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 667.00	1313.87	6.2e-64	1429	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 659.50	1306.21	1.6e-63	747	BL144664 602590932P1 NCI.CGAP.1
gb_est2:BG569281	+ 657.00	1306.87	1.5e-63	455	BL144664 602590932P1 NCI.CGAP.1

gb_est2:BL144664 + 656.50 1300.70 3.3e-63 713 BL144664 602590932P1 NCI.CGA
 gb_est2:BG569281 + 649.00 1286.74 2.0e-62 645 BL144664 602590932P1 NCI.CGA
 gb_est1:BL144664 + 642.00 1276.83 7.1e-62 448 BL144664 602590932P1 NCI.CGA
 gb_est1:BG569281 + 637.00 1263.11 4.1e-61 615 BL144664 602590932P1 NCI.CGA
 gb_est1:AB050299 + 631.00 1252.95 1.5e-60 520 BL144664 602590932P1 NCI.CGA

seq_name: gb_hhc:AK004045

seq_documentation_block:
 LOCUS AK004045 1446 bp mRNA linear HTC 19-JAN-2002
 DEFINITION Mus musculus 18 days embryo whole body cDNA, RIKEN full-length
 enriched library, clone:1110032P04:carboxypeptidase B2 (plasma),
 full insert sequence.

ACCESSION AK004045 GI:12835067
 VERSION AK004045.1 GI:12835067
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,
 clone:1110032P04.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE
 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 11042159
 PUBMED 11042159

REFERENCE
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Katsunagi,T., Taahito,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wachihi,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

REFERENCE
 AUTHORS The RIKEN genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 MEDLINE 11076861

REFERENCE
 AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
 Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
 Carlandi,P., Fukuda,S., Fukunishi,Y., Furuno,M., Haneagaki,T.,
 Hara,A., Hayatsu,N., Hill,D., Himoto,K., Hirooka,T., Hoti,F.,
 Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
 Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
 Kurahara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
 Numata,K., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
 Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
 Schiml,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T.,
 Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
 Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamana,T.,
 Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M., and
 Hayashizaki,Y.

REFERENCE
 AUTHORS Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

Direct Submission
 Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),


```

1155 ANCTATGATTGGGCAATCAAAATATCGTTTACAAATGAGCTCCAGATAC 1204
324 .....ThSerAnProProval 329
1205 AGCGAGATACGAGATTCTGCTGCTGAGAGATACATCAAAACCACTTGTG 1254
330 GIUlySleuLeuProleuSerleuLys 338
1255 CAGAGCTTGGCGCGCATCTCTAATA 1281
seq_name: gb_est1.AL536069

seq_documentation block:
LOCUS AL536069 889 bp mRNA linear EST 13-FEB-2001
DEFINITION AL536069 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF022YE21 5
prime, mRNA sequence.
ACCESSION AL536069
VERSION AL536069.1 GI:12799562
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 889)
AUTHORS L4,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.
FEATURES
source
1.889
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF022YE21"
/clone_lib="LTI_FL013_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMWSPORT 6; 1st strand
cDNA was primed with a NotI-collg(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMWSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifestech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 255 a 186 c 192 g 255 t 1 others
ORIGIN
alignment_scores:
Quality: 1244.50 Length: 276
Ratio: 5.207 Gaps: 1
Percent Similarity: 86.594 Percent Identity: 86.232
alignment block:
US-09-980-881-3 x AL536069 ..
Align seg 1/1 to: AL536069 from: 1 to: 889
1 pGcgInserGlgInValleuAlaIaIeuprOargzThrSerArgGlnVa 17
|||||
57 TTTCAGAGTGGCCAAAGTCTTACGTGCTCTTCCCTAGAACCTTAGGCAACT 106
|||||
17 IGIInValleuGlnAsnIleuThrThrThyTgIuIleValleuTpgInP 34
|||||
107 TCAGAGTTTACAGAAATCTTACTACAAACATATGATGTTCCTGTGCACAC 156
|||||
34 rovalThralAspleuIleValIySlySlyGlnValHispeheval 50

```

157 CGGTAAACAGCTGACCTTATTGTGAAGAAAAAACCAAGTCCATTTTTTTGTGA 206

51 AenAlaSerAspValAlaAspAsnValLysAlaHisLeuAsnValSerGly11 67

207 AATGCATCTGATGTGCACAATGTGAAGGCCCAATTAAATGTGAGCGGAAT 256

67 eProCysSerValIleuLeuAlaAspValGluAspLeuIleGlnGlnG11 84

257 TCCATGCAATGTCTTCTGCGACAGCTGGAGAAGATCTTATTCAACACACAGA 306

84 IeSerAsnAspThrValSerProAlaAlaSerAlaSerTyrTyrGlnGln 100

307 TTTTCCAAACACACAGTCCAGCCCCGAGCCCTCCGCATCTGACTATGAAACAG 356

101 TyrHisSerLeuAsnGluTyrSerTyrPheGluPheIleThrGlu117

357 TATCACTCACTAAATGAATCTATTCTTGATATGAATTTATTAATCTAGAG 406

117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys134

407 GCATCCTGATATGCTTACAAAAATCCACATTTGATCTCATTTTGAAAGAT 456

134 yrrProLeuTyrValIleuLysValSerGlyLysGlnThrAlaLysAsn 150

457 ACCCACTCTATGTTTAAAGTTCTTGAGAAACAAACACAGCCAAAGAT 506

151 AlaIleTyrIleAspCysGlyIleHisAlaArgGluTyrIleSerProAl 167

507 GCCATATGATTTGACTGTGGAAATCCATGCCAGAAATGATCTCTCTGCG 556

167 apheCysLeuTrrPheIleGlyHis 175

557 TTTCTCTGTGTGGTTCATAGCCCATATTAATCTATGGATTAATAG 606

175 175

607 GGCAATATACCAATCTCTGAGGCTTGGGATTTCTATGTTATGCCGGTG 656

176 AsnArgMetTrrPAr 180

657 GTTAATGTGATGGATTATGACTACTCATCATGAAAGAAATCGAATGTGAG 706

180 gLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuA 197

707 AAGAAGACCGTCTCTTCTATGCGAACATATATTCATATGGAGAACGACCTGA 756

197 snArgAsnHeAlaSerLysHisTrrPcysGlyGluGluLysAlaSerSer 213

757 ATAGGAACTTGTCTTCCAAACACTGTGTGAGAGAAAGTGCATCCACTTTC 806

214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVa 230

807 TCATGCTCGGAAACCTACTGTGACTTATCCGACTCAGAACCAAGAG 856

230 IlysAlaValAlaSerPheLeuArgTyr 239

857 GAAGGCAATGGCTAGTTCTTGTGAAGA 884

seq_name: gb_est2:B1144644

seq_documentation_block: 853 bp mRNA linear EST 05-JUL-2001

LOCUS B1144644 853 bp mRNA linear EST 05-JUL-2001

DEFINITION 60290939JF1 NCL_CGAP_L19 Mus musculus cDNA clone IMAGE:5051358 5',

ACCESSION mRNA sequence.

VERSION B1144644

KEYWORDS B1144644.1 GI:14604645

SOURCE EST.

ORGANISM house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 853)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey F. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: L1M1139 row: 0 column: 07
 High quality sequence stop: 849.

FEATURES
 source
 Location/Qualifiers
 1..853
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_image="501358"
 /clone_id="NCL_CGAP_L19"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: Liver; Vector: pCMV-SPORE; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.9 kb. Constructed by Life Technologies. Note: This is a NCL_CGAP Library."

BASE COUNT 243 a 199 c 194 g 217 t
 ORIGIN

Alignment_scores:
 Quality: 988.00 Length: 267
 Ratio: 4.638 Gaps: 4
 Percent Similarity: 79.775 Percent Identity: 72.659

Alignment_block:
 US-09-980-881-3 x B1144644 ..

Align seg 1/1 to: B1144644 from: 1 to: 853

```

1  PheGlnSerGlyGlnValIleuAlaIleuProArgThrSerArgGlnVa 17
   |||||
59  TTTCAGAGTGGCGAGTTTATCTCTCTCCAGAACCTCCAGGCAAGT 108
   |||||
17  GlnValIleuGlnAsnLeuThrThrTyrGluIleValIleuThrGlnP 34
   |||||
109  TCACCTACTCTCAGAACTTCTACTACAACTGAGGCTCTCTCGGCGAG 158
   |||||
34  rovalIthAlaAspLeuIleValIleuValIleuValIleuValIleu 50
   |||||
159  CAGTCAGACGCTCAATTCATCCAGAAAGAAAGAAAGTCCACTTTTGTG 208
   |||||
51  AsnAlaSerPValIleuValIleuValIleuValIleuValIleuVal 67
   |||||
209  AATCGCTGTGATGTCAGACAGTCAAGCGCATTAATGAGAGAGAT 258
   |||||
67  eProCysSerValIleuValIleuValIleuValIleuValIleuVal 84
   |||||
259  TCCATTAACTCTGATGAGAACAGTGGAGACCTAATTAACAGCAGA 308
   |||||
84  leSerAsnAPthValSerProArgAlaSerAlaSerAlaSerTyrGluGln 100
   |||||
309  CTTTCATGACAGCGTCAGCCCGCGCTCCCTTCACTACTATGAGCAG 358
   |||||
101  TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluAr 117
   |||||
359  TATCACTCGCTAAATGAATCTATCTCGATAGAGTCAAGTCAACTGAACA 408
   |||||
117  gHisProAspMetLeuThrIleHisIleGlySerSerPheGluYust 134
   |||||
409  GCATCTGACATGCTCCAGAAATCTACATCGATCATTCATTCAGAAAGT 458
   |||||
134  yrrProLeuTyrValIleuValIleuValIleuValIleuValIleuVal 150

```

seq_name: gb_es12:BG569281

seq_documentation_block:

LOCUS BG569281 852 bp mRNA linear EST 10-APR-2001
 DEFINITION 602588573F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722354 5',

ACCESSION BG569281
 VERSION BG569281.1 GI:13576934

KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 852)

REFERENCE
 AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 CDNA Library Preparation: CLONETECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: L1M1580 row: b column: 19
 High quality sequence stop: 751.

FEATURES
 source

Location/Qualifiers
 1..852
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4722354"
 /clone_id="NIH_MGC_76"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: Liver; Vector: pMDR-LIB (Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcgcttggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAAGGCGGAGGCGCGACACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 251 a 177 c 187 g 237 t
ORIGIN

alignment_scores:
Quality: 978.00 Length: 260
Ratio: 4.866 Gaps: 5
Percent Similarity: 77.308 Percent Identity: 75.385

alignment_block:
US-09-980-881-3 x BG569281 ..

Align seg 1/1 to: BG569281 from: 1 to: 852

```

135 ProleuTyrtAlleuysValserGlyLysgluThrAlaValAsnAl 151
|||||
2 CCACCTCATGCTTTTAAAGGTTCTGGAAAGAACAGACCCAAATGTC 51
151 alleTTPleAspCysGlylleHsAlaArgLutPrlSerProAlaP 168
|||||
52 CATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 101
168 heCysleuTrpHeileGlyHis..... 175
102 TCTGCTTGTTGTCATAGCCATATACCAATTCATGGAATATAGG 151
175 ..... 175
152 CAATATACCAATCTCTGAGGCTTGAGATTCATGTTATGCCAGTGT 201
176 .....AsnArgMetTrp.Arg 180
202 TAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 251
181 LysAsnArgSerPheTyrtAlaAsnAsnHisCyslleGlyThrAspLeuAs 197
252 AAGAACCGTCTTTCATGCGAACCAATCATGTCATCGAACAGACCTGAA 301
197 nAGAsnPhAlaSerLysHisTrpCysgluGluGlyAlaSerSerSers 214
302 TAGAACCTTGCTCCAAACAGTGGTGTAGAGGAGGCAATCCAGTCTCT 351
214 erCysSerGluThrTyrtCysGlyLeuTyrtProGluSerGluProGluVal 230
352 CATGCTCGGAAACCTACTGACTTATCTGATGATGATGATGATGATG 401
231 LysAlaValAlaSerPheLeuArgAsnIleAsnGlnIleLysAlaTy 247
402 AAGGAGAGTGGCTAGTTCTTGAGAGAAATATCAACAGATTAAGCATA 451
247 rIleSerMetHisSerTyrtSerGlnHisIleValPheProTyrtSerTyrt 264
452 CATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATATATTCCTATA 501
264 hrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280
502 CAGAGAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGTGA 551
281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrtHisGln 297
552 GCAGTTCGCTCTATTCAGAAAATTAGTAAATAATACAGTATACATGTC 601
297 yHisGlySerGluThrLeuTyrtLeuAlaProGlyGlyLysAspSerTrp 314
602 GCATGGCTCGAAACCTTATACCTAGCTCTGGAGGTGGGAGCATTTGA 651
314 leTyrtAspLeuGlylleLysTyrtSerPheThr..... 324
652 TCATATGATTTGGGACATACATATATGCTTTACCAATGACCTTCGAGATA 701

```

```

325 .....SerAsn..ProProva 329
702 CGGCGACATTAACGAGTTCCTGCTCCGAGCGGTACATCAACACCTGT 751
329 lGluLysleuLeuProLeuSerLeu 337
752 AGAGAGCTTTTGCGGCTGTCTTA 776

```

seq_name: gb_est2:BG616456

seq_documentation_block: 738 bp mRNA linear EST 18-APR-2001

DEFINITION 602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:473500 5', mRNA sequence.

ACCESSION BG616456

VERSION BG616456.1 GI:13667827

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgi.ncl.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Email: cgrabbs-remail.nih.gov

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

CDNA Library Preparation: CLONTECH Laboratories, Inc.

Tissue Procurement: CLONTECH Laboratories, Inc.

CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LLCM1595 row: c column: 05

High quality sequence stop: 666.

Location/Qualifiers

1..738

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:473500"

/clone_1lb="NIH_MGC_76"

/lab_host="DH10B (TI phage-resistant)"

/note="Organ: liver; Vector: pDRR-LIB (Clontech); site_1: 5' flit (ggcgctcgcc); site_2: 3' flit (ggcattatggc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTAAGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGCGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 209 a 161 c 153 g 215 t
ORIGIN

alignment_scores:
Quality: 926.50 Length: 283
Ratio: 4.391 Gaps: 4
Percent Similarity: 74.558 Percent Identity: 67.845

alignment_block:
US-09-980-881-3 x BG616456 ..

Align seg 1/1 to: BG616456 from: 1 to: 738

```

1 PhgGlnSerGlyGlnValAlaLeuProArgTrpThrSerArgGlnVal 17
|||||
40 TTTCAGAGTGGCCAAAGTTCTAGCTCTTCTTCAAGCACTTGAAGT 89
17 lGlnValleuGlnAsnleuThrTrpTrpGlnIleValleuTrpGlnP 34
|||||

```

```

90 TCAAGTCTACAGATCTTACTACAAATATGAGATGTTCTCTGCAGC 139
34 rovalThralaspLeuIleValLysLysGlnValHisPhePheVal 50
140 CGGTAAACAGCTGACCTTATGTGTAAGAAAAACAAGCTCATTTTGTGA 189
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyI 67
190 AATGCATCTGATGTCGACAAATGGAAGCCCAATTAATGTCGAGCAAT 239
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
240 TCCATCAGTGTCTTCTGTCGACAGCTGGAAGATCTTATTAACACACAG 289
84 leSerAsnAspThrValSerProAlaGalaSerAlaSerTyrGluGln 100
290 TTTCCAAACACACAGTCAGCCCCGAGCCCTCCGATCTACTATGAACAG 339
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIlePheGlu 117
340 TATCACTCTCAATGAATGAATCTATCTTGATGATGAATTAATTAACAG 389
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
390 GCATCCTGATGCTTACAAAATCCACATGATGATCCATTTGAGAACT 439
134 yrrProLeuTyrValLeuLysValSerGlyLysGlnIleThrAlaLys 150
440 ACCCACTCTATGTTTAAAGGTTTCTGGAAGAAACACACCAAAAT 489
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerPro 167
490 GCCATATGATGATGACTGCGAATCCATGCCAGAGATGATCTCTCC 539
167 aPheCysLeuTrpPheIleGlyHis..AsnArgMetTrpArgLysAsn 183
540 TTTCTGCTGTGTCATAGGCCATATACATCAATTCATGGAATAATG 589
183 gSerHeTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsn 200
590 GGCATATATACCAATCTCTGAGCTGTGGA..... 620
200 heAlaSerLysHisTrpCysGluGlnGlyAlaSerSerSerCysSer 216
621 .....TTTCTATGTTATGCCGGGTTC..... 644
217 GluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 233
645 .....TGTGACTG..... 653
233 lAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSer 250
653 ..... 653
250 eHisSerTyrSerGlnHisIleValPheProTyrSerTyrThrArgSer 266
654 .....GTTATGACTACTCATGCGCAAGAT 680
267 LysSerLysAspHisGluGlnLeuSerLeuValAlaSerGluAlaVal 282
681 CGCACTGTTGGACAAAGAACCCGCTTACTAATGCAAAACAAT 728

seq_name: gb_est2:BI332659
seq_documentation_block:
LOCUS BI332659 1022 bp mRNA linear EST 30-JUL-2001
DEFINITION 602984408F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5137369 5',
ACCESSION BI332659
VERSION BI332659.1 GI:15017316
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LRAM1337 row: 0 column: 02
High quality sequence stop: 841.

FEATURES
Source
Location/Qualifiers
1..1022
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5137369"
/clone_1ib="NCI_CGAP_L19"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: Liver; Vector: PCMV-SpOrf6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 303 a 233 c 225 g 261 t
ORIGIN

alignment_scores:
Quality: 924.00 Length: 316
Ratio: 3.949 Gaps: 6
Percent Similarity: 74.051 Percent Identity: 61.709

alignment_block:
US-09-980-881-3 x BI332659 ..

Align seg 1/1 to: BI332659 from: 1 to: 1022

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnVal 17
|||||
60 TTTCAAGTGGCCAGCTTTATCTGCTCTCCAAAGACCTCAGCAGCAAGT 109
17 GlnValLeuGlnAsnLeuThrThrTyrGluIleValLeuTyrGlnP 34
|||||
110 TCACTACTTCAGAACTTACTACAAAGATGAGGTCGTCTCTGCGACG 159
34 rovalThralaspLeuIleValLysLysGlnValHisPhePheVal 50
|||||
160 CAGTGACAGCTGAATTCATCGAAGAAAGAAAGAGCTCAGCTTTTGTG 209
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyI 67
|||||
210 AATGCGTGTGATGTCGACAGTCTCAAGCGCATTTAAATGTGAGCAGAT 259
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
|||||
260 TCCATTTAAGCTTGTGATGAACACAGCTGAGGACCTTAATTAAACAGCA 309
84 leSerAsnAspThrValSerProAlaGalaSerAlaSerTyrGluGln 100
|||||
310 CTTTCAATGACACAGCTGAGCCCGCGCTCCTCATTAAGAGAGAG 359
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 117
|||||
360 TATCACTGCTAAATGAATCTATCTCGATAGAAAGTCATCACTAGCAACA 409
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
|||||

```

410 GCATCCTGCATGCTCCAGAAAATCTACATCGCATCTCTTGAGAACT 459
 134 yPProLeuTyValLeuLysValSerGlyLysGluInThrAlaLysAsn 150
 460 ACCACCTTTATGTTTAAAGCTCTCAGGAAAGACAAAGAACCAAAAT 509
 151 AlaIleTrpIleAspCysGlyIleHisAlaArgLutTrpIleSerProAl 167
 510 GCCATCTGATGCTGATGGAATCCATGCCAGAAATGATTTCCACTGC 559
 167 aPheCysLeuTrpPheIleGlyHis 175
 560 TTTCCTGTTC TGCTTCATAGGCTACGTGACACATTCACATGCCGAT 608
 175 175
 609 ATCTGATACCAAGACTTCTGAGGACGTGGAATTTCTACATCATGCCGT 658
 176 176
 659 GATGGAACGTGATGCTATGCTATGACTACACGTGACAAAAGAAATGGAATGTT 708
 179 pArgLysAsnArgSerPheTyAlaAsnAsnHisCysIleGlyThr...A 195
 709 GCAGGAAGAACCGTCTGCTCACCAGAACACCGCTGCTGTGGCAGACATG 758
 195 sPheAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSer 211
 759 ACCTGAACAGAACTTCGCTCCAAACACGTGTGTGAAAGGTGGCTC 808
 212 SerSerSerCysSerGluThrTrpTyCysGlyLeuTyP ProGluSer Glu 227
 809 AAGTTTCCTCCTCTCTGAAACTGCTGACCTTATTCCTGAGCTGTGAG 858
 228 ProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlu 244
 859 CCAGAACTGAAGCCAGGCGGTGACTTTTGAAGAAGATATCTGACCCCT 908
 244 eLysAlaTyrlIleSerMetHisSerTySerGlnHisIleValPhePro. 260
 909 TAAAGTT.....ACTTCAGATGGCTCATATATCTCCAC 940
 261TySerTyTrpArgSerLysSerLysAspHisGlu 272
 941 AATATGTTCTCTATTTCATATGAAAGACGAGACCAAGAA 984

seq_name: gb_est2:BG618629

seq_documentation_block:
 LOCUS BG618629 788 bp mRNA linear EST 18-APR-2001
 DEFINITION 602646064F1 NIH-MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',
 mRNA sequence.
 ACCESSION BG618629
 VERSION BG618629.1 GI:13670000
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 788)
 NIH-MGC http://mhc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM1630 Row: 0 Column: 15

FEATURES
 source High quality sequence stop: 760.
 Location/Qualifiers
 1..788
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4767974"
 /clone_11b="NIH-MGC_76"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Liver; Vector: pDMR-LIB (Clontech); Site_1:
 SfiI (ggcgccctcgcc); Site_2: SfiI (ggccatctggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCCGAGCGCGCAGCATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH-MGC Library."

BASE COUNT 222 a 157 c 169 g 240 t
 ORIGIN

alignment_scores:
 Quality: 921.00 Length: 219
 Ratio: 5.088 Gaps: 1
 Percent Similarity: 82.648 Percent Identity: 81.735

alignment_block:
 US-09-980-881-3 x BG618629 ..

Align seg 1/1 to: BG618629 from: 1 to: 788

1 PheGlnSerGlyGluValLeuAlaAlaLeuProArgThrSerArgLysVal 17
 132 TTTGAGAGTGGCCAAAGTTTACGTAGCTGCTCTCTGAAACCTCTAGCAGT 181
 17 LglnValLeuGlnAsnLeuThrThrTrpGluIleValLeuTrpLys 34
 182 TCAAGTTCTACAGAACTTCTACACATATGAGATGTTCTCTGCGAGC 231
 34 roValThrAlaAspLeuIleValLysLysGluValHisPhePheVal 50
 232 CGGTACACGCTGACCTTATTTGTGAAGAAAAACAAGTCCATTTTGTGTA 281
 51 AsnAlaSerAspValAlaAspAsnValLysAlaHisLeuAsnValSerGly 67
 282 AATGATCTGATGTGCAAAATGTGAAGCCCATTTAATGTGACGGAGAT 331
 67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 84
 332 TCCATGCAAGTGTCTTGGTGGCAGATGTGAAGATCTTATTCACACAGAG 381
 84 leSerAsnAspThrValSerProArgAlaSerAlaSerTyTrpGluGln 100
 382 TTTCACACGACACAGTCAGCCCGAGCGCTCCGATGCTATATGAAACAG 431
 101 TyrHisSerLeuAsnGluIleTySerTrpIleGluPheIleThrGluArg 117
 432 TATACACACAAATGAATCTATCTTGGATGAATATTTAATGAGAG 481
 117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
 482 GCATCTGATATGCTTACAAAATCCACATGATGATCCCATTTGAGAGT 531
 134 yPProLeuTyValLeuLysValSerGlyLysGluInThrAlaLysAsn 150
 532 ACCACCTTATGTTTAAAGCTTCTGAGAAAGACCAAGCCAAAT 581
 151 AlaIleTrpIleAspCysGlyIleHisAlaArgLutTrpIleSerProAl 167
 582 GCCATATGATGATGCTGGAATCCATGCCAGAAATGATGATCTCTGCG 631
 167 aPheCysLeuTrpPheIleGlyHis..... 175


```
|||||
632 TTTCTCTGTGTTGCTCATAGGCCATATMACTCATTTCTATGGATATAG 681
175 .....
682 GGCAATATTACCAATCTCTGAGGCTTGATTTCTATGTATGCGAGT 731
176 .....
732 GGTAAATGTTGATGGTTATGACTACTCATGAGAAAGAAATCGATTGTG 781
179 PARGLYS 181
|||||
782 GAGAAAG 788

seq_name: gb_est2:BG565100

seq_documentation_block:
LOCUS BG565100 747 bp mRNA linear EST 10-APR-2001
DEFINITION 602583720P1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4711348 5',
mRNA sequence.
ACCESSION BG565100
VERSION BG565100.1 GI:13572753
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 747)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: egads+remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1CM1551 row: h column: 05
High quality sequence stop: 714.
Location/Qualifiers
1. 747
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4711348"
/clone_1kb="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
5' (ggcgccgcgcgc); Site_2: 5' (ggccatagcgc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGAGCGCGGCAGATC-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 220 a 154 c 166 g 207 t
ORIGIN
```

```
alignment_scores:
Quality: 912.00 Length: 251
Ratio: 4.449 Gaps: 6
Percent Similarity: 81.673 Percent Identity: 78.088
```

```
alignment_block:
US-09-980-881-3 x BG565100 ..
```

```
Align seq 1/1 to: BG565100 from: 1 to: 747
```

```
30 ValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGlnVal 46
|||||
1 GTTCTGTGCGAGCGGTAAACAGTGCCTTAT. GTCAAAAAAACCAAGT 49
|||||
46 HisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisLeu 63
|||||
50 CCATTTTGTGATGATGATCATGATGTCGACAAATGAAAGCCCATTTTAA 99
|||||
63 snValSerGlyIleProCysSerValLeuLeuAlaAspValGlnAspLeu 79
|||||
100 ATGTAGCGGAATTCATCATGCTGCTTCTGCGAGATGGAAGATCTT 149
|||||
80 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAla 96
|||||
150 ATTCAACACAGCATTTCCAAACGACAGTCAGCCCCGAGCCTCCGCATC 199
|||||
96 TTTTGTGTGAGTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 249
|||||
200 GTACTATGACAGTATCATCTCACTAAATGAAATCTATTCTTGATGAT 249
|||||
113 heIleThrGlnArgHisProAspMetIleThrLysIleHisIleGlySer 129
|||||
250 TTATTAACGTAGAGGATCTCTGATATGCTTACAAAATCCACATTTGATCC 299
|||||
130 SerPheGlnLysTyrProLeuTyrValLeuLysValSerGlyLysGln 146
|||||
300 TCATTTGAGAACTACCACTATGCTTTTAAAGTTTCGAAAGAACAA 349
|||||
146 nThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgLys 163
|||||
350 AGCAGCCAAAATGCGCATATGATGATGATGATGATGATGATGATGAT 399
|||||
163 rPheSerProAlaPheCysLeuTrpPheIleGlyHis..... 175
|||||
400 GGACCTCTCCCTCTTCTGCTTGTGTTGATAGGCCATTAATCAATTC 449
|||||
175 ..... 175
|||||
450 TATGGATTAATAGGCAATATACCAATCTCGAGGCTTGATTTCTA 499
|||||
176 ..... 180
|||||
500 TGTATGCCAGTGTAAATGTGATGTTATGACTCATGAGAAACAG 549
|||||
181 .....LysAsnArgSerPheTyrAlaAsnHisCys 191
|||||
550 AATCGAATGTGAGAACAGAACGCTTCTTCTATGCGAACATCATTC 599
|||||
192 IleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTyrCysGln 208
|||||
600 ATCGGAACAGACCTGAATAGGAACCTT. GCTTCCAAACACTGGTGTGAGA 648
|||||
208 u.GlyAlaSerSerSerSerCysSerLysThrTyrCysGlyLeuTyr 223
|||||
649 AGGTGATCCAGTCTCTCATGCTCGGAAAACTACTGTTGAGCTTAT 698
|||||
224 ProGlnSerGlnProGlnValLysAlaValAlaSerPheLeuArg 238
|||||
699 COTGAGTCAGAACGAGAGTG. GAGGACATGGCTACTTCTTGAAAG 742

seq_name: gb_est2:BG566051

seq_documentation_block:
LOCUS BG566051 747 bp mRNA linear EST 10-APR-2001
DEFINITION 602583720P1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4710564 5',
mRNA sequence.
ACCESSION BG566051
VERSION BG566051.1 GI:13573704
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```


REFERENCE 1 (bases 1 to 747)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapds-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>
 Plate: L16M1549 row: 9 column: 13
 High quality sequence stop: 731.

FEATURES
 source location/Qualifiers
 1. 747

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4710564"
 /clone_lib="NIH_MGC_76"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgccggccg); Site_2: SfiI (ggcgatcgccg); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCATATGACC-3' and 3' adaptor sequence: 5'-ATTCAGAGCGCCGAGCGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 212 a 157 c 159 g 219 t
 ORIGIN
 alignment_scores
 Quality: 897.00 Length: 195
 Ratio: 4.849 Gaps: 2
 Percent Similarity: 94.872 Percent Identity: 91.282

alignment_block:
 US-09-980-881-3 x BG566051 ..

Align seg 1/1 to: BG566051 from: 1 to: 747

```

1 pheglnserc]glnvalleualaaleuProArGThSerarglnva 17
|||||
86 TTTCAGAGTGGCCAACTTACGCTGCTCCATTAACCTGAGCAAGT 135
|||||
17 lgnvalleuglnasnleuThrThrThrYrgluilevalleuTrpInp 34
|||||
136 TCAGAGTTTACAGAACTTACTACACATATGAGATTGTTCTGCGACG 185
|||||
34 rovalThraIaAspleuilevallylslysglnvalhisphelval 50
|||||
186 CGGTACAGCTGACCTTATTGTGAGAAAAACAGACCATTTTGTGTA 235
|||||
51 AsnAlaSerAspValaspasnvallylsalhisleuasnvalsergyl 67
|||||
236 AATGATCTGATGTCACAAATGTGAAGCCCATTTAAATGTGAGCGAAT 285
|||||
67 eProCySerValleuLeuAlaAspValgluAspleuileglnInp 84
|||||
286 TCATATCAGAGTCTTCTGTCGACACAGTGGAGAAATCTTATTCAACAGCAGA 335
|||||
84 lSerAsnAspThrValSerProArGAlaSerAlaSerYrYrgluIn 100
|||||
336 TTTCACAGCAGACAGTACAGCCCGGAGCCCTCCGATCTACTATGAACAG 385
|||||
101 TyrHisSerleuasngluileYrSerTrpIlegluPhelIleThrgluar 117
|||||

```

```

386 TATCACTCACTAAATGAATCTATTCTTGATAGCAATTTTAACAGAG 435
|||||
117 ghisProAspMetleuThrlysilehisileglyserSerPhegluYst 134
|||||
436 GCATCCGATATGCTTACAAAATCCACATTTGATCTCTATTGAGAGT 485
|||||
134 YrProLeuYrVal.leuYsValSerGlylysgluInThraIaIyas 150
|||||
486 ACCACACTCTATGTTCTTAAAGTTTCGAAAAGAACACAGCAGCAAAA 535
|||||
150 naIaIeTrpIlaAspCysglyllehisAlaArggluTrpIleSerPro 167
|||||
536 TGGCAATATGATGACTGTGAATCCATGCCAGAGAAATGATCTCTCG 585
|||||
167 laphCysleuTrpPheIleGlyHis..AsnArgMetTrpArgLysAsn 183
|||||
586 CTTTCTGCTTGTGGTTTCATATGAGCATTAATCTCATCTATGGATATA 635
|||||
183 rSerPheYrAlaAsnAsnHisCysIleGly 193
|||||
636 GGGCAATATACCAATCTCTGAGGCTTGGA 667
|||||

```

seq_name: gp_est1:AI048167
 seq_documentation_block: 724 bp mRNA linear EST 08-JUL-1998
 LOCUS AI048167
 DEFINITION ud74b12.y1 Sugano mouse liver mla Mus musculus cDNA clone
 IMAGE:1451327 5' similar to TR:015114 Q15114 PREPRO-PLASMA
 CARBOXYPEPTIDASE B. ;, mRNA sequence.
 ACCESSION AI048167
 VERSION AI048167.1 GI:3296454
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 724)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HM Mouse EST Project
 Unpublished (1996)
 CONTACT: Marra M/Mouse EST Project
 WASHU-HM Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNLN ; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:924643
 Possible reversed clone: similarity on wrong strand
 Seq primer: custom primer used
 High quality sequence stop: 405.

FEATURES
 source

1. 724
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:1451327"
 /clone_lib="Sugano mouse liver mla"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: liver; Vector: pME185-FL3; Site_1: DraIII
 (CACTGTG); Site_2: DraIII (CACCAGTG); 1st strand cDNA
 was primed with an oligo(dT) primer
 [ATGAGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was
 ligated to a DraIII adaptor [TCTTGGCTACTAGG], digested
 and cloned into distinct DraIII sites of the pME185-FL3

vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCTCTTAAAGCTGGC and 3' end primer CGACCTCGACCTGACACA."

BASE COUNT 215 a 169 c 160 g 178 t 2 others
ORIGIN

alignment_scores:
Quality: 894.50 Length: 241
Ratio: 4.639 Gaps: 1
Percent similarity: 79.668 Percent identity: 70.954

alignment_block:
us-09-980-881-3 x AT048167 ..

Align seg 1/1 to: AT048167 from: 1 to: 724

```

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnVa 17
|||||
2 TTTCAGAGTGGCCAGCTTTTATCTCTCTCCACAGACCTCCAGCAGACT 51
|||||
17 GlnValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrrpglnp 34
|||||
52 TCACACTACTCAGAACTTACTACAACTATGAGCTGCTCTCTGCGAGC 101
|||||
34 rovalttrAlaAspleuIleValLysLysGlnValHisphetheVal 50
|||||
102 CAGTGACAGCTGAATTCATCCAGAAAGAAAGAAAGTCCACTTTTGTG 151
|||||
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGly1 67
|||||
152 AATGGCTGTGATGTGACAGCTCAAGCGCATTTAAATGTGAGCAGAA 201
|||||
67 eProcysSerValLeuLeuAlaAspValGluAspleuIleGlnGlnInt 84
|||||
202 TCCATTATACGTTCTGATGACACAGTGGAGGACCTAATTGACAGCAGA 251
|||||
84 LeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGln 100
|||||
252 CTTTCAATGACACAGCTCAGCCCGCCCTCCCTTCACTATGACAGCAG 301
|||||
101 TyrHisSerLeuAsnGluIleTyrSerTrrpIleGlnpheIleThrGlu 117
|||||
302 TATCACTCGCTAAATGAATCTATTCCTGATAGAAAGTCATTAAGTACA 351
|||||
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
|||||
352 GCATCTGCTGACATGCTCCAGAAATCTACATCGATCATCTCGAAGT 401
|||||
134 YrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
|||||
402 ACCCACTTATGTTTAAAGCTCAGACAGAGAAACAAAGATCAAAAT 451
|||||
151 AlaIleTrrpLeaAspCysGlyIleHisAlaArgGluTrrpIleSerPro 167
|||||
452 GCCATCTGAGTGAAGTGAATCCATGCCAGAGAAATGATTTTCACTGCG 501
|||||
167 aPhcCysLeuTrrpPheIleGlyHis..... 175
|||||
502 TCTCTGTTGTGGTTCAATAGCTACGTGACACAAATTCATGAGAGAA 551
|||||
175 ..... 175
552 ATCTGTATACAGACTTCTGAGCAGCGATCTCTACATCATGACAGCTG 601
|||||
176 .....AsnArgMetLrrpAr 180
|||||
602 ATGAACGTGATGGCTATGACTACACGTGAAAAAGAAATGCAATGTGAG 651
|||||

```

180 gLysAsnArgSerPheTrrpAlaAsnHisCysIleGlyThrAspleuA 197
|||||
652 GAAGAGCCGNTCTGCTCACAGAACACACCGCTGGGACAGACCTGA 701
|||||
197 snArgAsnPhaAlaSerLysHis 204
|||||
702 ACAGGAACCTCGCTCCAAACAC 724
seq_name: gb_est1:AV646979

seq_documentation_block:

LOCUS AV646979 735 bp mRNA linear EST 15-JAN-2002
DEFINITION AV646979 GIC Homo sapiens cDNA clone GICATD06 3', mRNA sequence.
ACCESSION AV646979
VERSION AV646979.1 GI:9867993
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
1 (bases 1 to 735)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE
JOURNAL Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
MEDLINE Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
COMMENT Hu, G., Gu, J., Chen, Z., and Han, Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

CONTACT: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@cnsc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source location/Qualifiers
1. 735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GICATD06"
/clone_lib="GIC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 200 a 158 c 172 g 205 t
ORIGIN

alignment_scores:
Quality: 863.00 Length: 176
Ratio: 5.047 Gaps: 1
Percent similarity: 97.159 Percent identity: 96.591

alignment_block:
us-09-980-881-3 x AV646979 ..

Align seg 1/1 to: AV646979 from: 1 to: 735

```

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnVa 17
|||||
58 TTTCAGAGTGGCCAGCTTTTATCTCTCTCCACAGACCTCCAGCAGACT 107
|||||
17 GlnValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrrpglnp 34
|||||
108 TCAAGTCTCAGAACTTACTACACATATGAGATGTCTCTGCGAGC 157
|||||
34 rovalttrAlaAspleuIleValLysLysGlnValHisphetheVal 50
|||||

```

```

158 CGGTAACAGTGACCTATTGTGAGAAAAAACAGTCATTTTGTGA 207
51 AsnAlaSerAspValAspAsnValLysAlaHisLysAsnValSerGlyI 67
|||||
208 AATGATCTGATGTCGACAAATGTAAGCCCATTTAAATGTGAGCGAAT 257
67 eProcySerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnI 84
|||||
258 TCCATGCGACGTCTGCTGCGACAGCTGGAAGATCTTATTCAACACAGCA 307
84 LeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
|||||
308 TTTCACAGACACAGACAGACCCCGAGCCTCCGCATCGGACATGTAACAG 357
101 TyrHisSerLeuAsnGluIleTyrSerTyrIleGluPheIleThrGluar 117
358 TATCACTCACTAAATGAATCTATTCTTGATAGAAATTTATACTAGAG 407
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysT 134
|||||
408 GCATCCTGATGCTTACAAAAATCCACATGATCTCATTTGAGAGAT 457
134 TyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAs 150
|||||
458 ACCCACTCATATGTTTAAAGGCTGTGAAAAAGAACACACACCAAAAA 507
150 nAlaIleTyrPheAspCysGlyIleHisAlaArgGluTyrPheSerProa 167
|||||
508 TCCCAATATGATGATGCTGGAATCCATGCCAGAAATGATCTCTCCG 557
167 IapheCysLeuTyrPheIleGlyHis 175
|||||
558 CTTTCTGCTGTGGGGGTCAATAGGCCAT 583

seq_name: gb_est2:Bg618239

seq_documentation_block:
LOCUS Bg618239 725 bp mRNA linear EST 18-APR-2001
DEFINITION 602645967F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767376 5',
mRNA sequence.
ACCESSION Bg618239
VERSION Bg618239.1 GI:13669610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 725)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.llnl.gov
Plate: LCM1629 row: f column: 17
High quality sequence stop: 723.
Location/Qualifiers
1..725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4767376"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1:
SfiI (ggccgctcgcc); Site: 2: SfiI (ggccattatggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCATTTATGCG-3' and 3' adaptor sequence:

```

```

5'-ATCTAGAGCGCCGAGCGCGCAGATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 203 a 153 c 149 g 220 t
ORIGIN

alignment_scores:
Quality: 862.00 Length: 175
Ratio: 4.926 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.286

alignment_block:
US-09-980-881-3 x Bg618239 ..

Align seg 1/1 to: Bg618239 from: 1 to: 725

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlna 17
|||||
131 TTTCAGAGTGGCCAGTTCATAGCTCTTCTAGAACCTTAGGCGAGT 180
17 IglValLeuGlnAsnLeuThrThrTyrGluIleValLeuTyrGlnP 34
181 TCAAGTTCACAGAACTTCTACACATATGAGAT.GTTCCTGCGAGC 229
34 roValThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
230 CGGTACAGCTGACCTTAT.GTGAAGAAAAACAGTCCATTTTGTGA 278
51 AsnAlaSerAspValAspAsnValLysAlaHisLysAsnValSerGlyI 67
|||||
279 AATGATCTGATGTCGACAAATGTAAGCCCATTTAAATGTGAGCGGAAT 328
67 eProcySerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnI 84
|||||
329 TCCATGCGACGTCTGCTGCGACAGATGGAAGATCTTATTCAACACAGCA 378
84 LeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
|||||
379 TTTCACAGACACAGTACGCCCGAGCCTCCGCATCGTACTATGACAG 428
101 TyrHisSerLeuAsnGluIleTyrSerTyrIleGluPheIleThrGluar 117
|||||
429 TATCACTCACTAAATGAATCTATTCTTGATAGAAATTTATACTAGAG 478
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysT 134
|||||
479 GCATCCTGATGCTTACAAAAATCCACATGATCTCATTTGAGAGAT 528
134 TyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
|||||
529 ACCCACTCATATGTTTAAAGGCTGTGAAAAAGAACACACACCAAAAT 578
151 AlaIleTyrPheAspCysGlyIleHisAlaArgGluTyrPheSerProa 167
|||||
579 GCATATGATGATGCTGGAATCCATGCCAGAAATGATCTCTCTCC 628
167 apheCysLeuTyrPheIleGlyHis 175
|||||
629 TTTCCTGCTGTGGGGGTCAATAGGCCAT 653

seq_name: gb_est1:AV651709

seq_documentation_block:
LOCUS AV651709 688 bp mRNA linear EST 15-JAN-2002
DEFINITION AV651709 GIC Homo sapiens cDNA clone GLCSCF10 3', mRNA sequence.
ACCESSION AV651709
VERSION AV651709.1 GI:9872723
KEYWORDS EST.
SOURCE human.

```


Align seg 1/1 to: BG618813 from: 1 to: 806

```

178 MetTPARGlySAsnArgSerPheTYrAlaAsnAsnHisCysIleGly 194
    |||||||
2  ATGTGAGAAAGAACGCTTCTTCATGGCAACATCATTCATCGGAAAC 51
    |||||||
194 rAspleuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAla 211
    |||||||
52 AGACCTGAATAGAACTTGTCTTCCAAACACGTGCTGAGGAAGTCCAT 101
    |||||||
211 eSerSerSerCysSerGluThrTrpCysGlyLeuTyProGluSerGlu 227
    |||||||
102 CCAATTCCATGCTCTGGAACCTACTGTGACTTATCTCTGAGTCAGAA 151
    |||||||
228 ProGluValLysAlaValAlaSerPheLeuArgArgAsnIleasnGln 244
    |||||||
152 CCAGAGTAAAGGCACTGGCTAGTCTTCTTGAGAAAGAAATATCAACAG 201
    |||||||
244 eLysAlaTYrLleSerMetHisSerTYrSerGlnHisIleValPhePro 261
    |||||||
202 TAAAGCATATCATCAGCATCATCATCATCCACGATATAGTCTTCCAT 251
    |||||||
261 yPSerTYrThrArgSerLysSerLysAspHisGluGluLeuSerLeuVal 277
    |||||||
252 ATTCCTATACAGCAAGTAAAGCAAGAACATAGAGAACTGTCTAGTA 301
    |||||||
278 AlaSerGluAlaValAlaArgAlaIleGluLysThrSerLysAsnThrArg 294
    |||||||
302 GCACAGTAAGCACTGTCTGCTATTTGAGAAACAGTAAATAATCAAGTA 351
    |||||||
294 rThrHisGlyHisGlySerGluThrLeuTYrLeuAlaProGlyGlyAla 311
    |||||||
352 TACACATGGCGCATGCTCAGAAACCTTATCTACGCTCCGAGAGTGGGG 401
    |||||||
311 sPaSprPrlLeuTYrAspLeuGlyLleLysTYrSer.Phe..... 323
    |||||||
402 ACCATTGGATCTATGATTGGGCATCAATATTCGTTTCAATTAACACTT 451
    |||||||
324 .....ThrSerAsp 327
    |||||||
452 CGAGATACGGGCATACGATCTTCTGCGCGAGGCTTACATCAACAAAC 501
    |||||||
327 rProValGluLysLeuLeuProLeuSerLeuLys 338
    |||||||
502 CACCTGTAGAGAGCTTTGCCCTGCTCTCTTAA 536
    |||||||
seq_name: gb_est2:BG618813

seq_documentation_block:
LOCUS      BG618813              750 bp      mRNA      linear      EST 18-APR-2001
DEFINITION 602646186F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767885 5',
            mRNA sequence.
ACCESSION  BG618813
VERSION    BG618813.1 GI:13670184
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 750)
            NIH-MGC http://mgc.ncl.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/MLN at:
            http://image.llnl.gov

```

Plate: L16M1630 row: k column: 22
High quality sequence: 622.
Location/Qualifiers

FEATURES
source

1. 750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4767885"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site: 1;
SfiI (ggccattagcgc); Site 2: SfiI (ggccattagcgc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCGCATTTGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCCAGGCGCGCAGCATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 219 a 146 c 171 g 213 t 1 others
ORIGIN

alignment_scores:
Quality: 814.50 Length: 188
Ratio: 4.735 Gaps: 5
Percent Similarity: 91.489 Percent Identity: 89.362

alignment_block:
US-09-980-881-3 x BG618813 ..

Align seg 1/1 to: BG618813 from: 1 to: 750

```

1  PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnVal 17
    |||||||
77  TTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCCTAGAACCTTAGGCAAGT 126
    |||||||
17  LglnValLeuGlnAsnLeuThrThrTYrGluLleValLeuTrpGlnP 34
    |||||||
127  TCAAGTTCCTACAGAACTTACTACAAACATATGAGATGTTCTGCGCAGC 176
    |||||||
34  rovalThrAlaAspLeuLleValLysLysGlnValHisPhePheVal 50
    |||||||
177  CGGTAACAGTGCACCTTATTTGAGAAAGAAACAGTCCATTTTGTGTA 226
    |||||||
51  AsnAlaSerAspValaAspAsnValLysAlaHisLeuAsnValSerGly 67
    |||||||
227  AATGCATCTGATGCGACATGTGAAGCCCATTTAATGTGACGGGAAT 276
    |||||||
67  eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
    |||||||
277  TCCATGCGAGTGTCTTCTGCGCAGACGTGAAAGATCTTATTCACACAGCA 326
    |||||||
84  leSerAsnAspThrValSerProArgAlaSerAlaSerTYrTYrGluGln 100
    |||||||
327  TTTCACAGACACAGTCAGAACCCGAGCCCTCCGATGCTATGAGACAG 376
    |||||||
101  TYrHisSerLeuAsnGluLleTYrSerTrpIleGluPheIleThrGluAr 117
    |||||||
377  TATACATCATTAATGAATATATCTTGGATAGAAATTTAATGATGAGAG 426
    |||||||
117  gHisProAspMet.LeuThrLysIleHisIleGly.SerSerPheGluL 133
    |||||||
427  GCATCCGTGATATGCTTAAACAAAATCCACATTTGCAATCCATATCGAGAA 476
    |||||||
133  sTYrProLeu.TyrValLeuLysValSer.GlyLysGluGlnTrpAla 149
    |||||||
477  GTAACAGTGTATGTTTAAAGGTTCTCTGGGAAAGAAACAGAGGCA 526
    |||||||
149  sAsnAlaIleTrpLleAspCysGlyLleHisAlaArgGluTrpLleSerP 166
    |||||||
527  AATATGCATATGATTCATCTGTGAATCCATGCCAGAGATGATCTCTC 576
    |||||||

```

166 roAlaPheCysLeuTrp.....PheIleGlyHis 175
|||||
577 CTGCTTCTGCTGTGTGCATAGGCCCATATAACTCAAAATTCTATGGGAT 626
176 AsnArg 177
|||||
627 AATAGG 632

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:47:50 ; Search time 23.83 Seconds

(without alignments)
399,707 Million cell updates/sec

Title: US-09-980-881-4

Sequence: 1 ASASTYEQYHSLNEITSWIE.....IKSFTSNPPEKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	601	44.9	415	1 CBPB_RAT	P19223 rattus norv
2	586	43.8	306	1 CBPB_BOVIN	P00732 bos taurus
3	578	43.2	416	1 CBPB_CANFA	P55261 canis famli
4	575	43.0	401	1 CBPB_PIG	P09955 sus scrofa
5	567	42.4	417	1 CBPB_HUMAN	P15086 homo sapien
6	561	41.9	417	1 CBPC_MOUSE	P15089 mus musculi
7	552	41.3	309	1 CBPC_RAT	P12961 rattus norv
8	552	41.3	417	1 CBPC_HUMAN	P15088 homo sapien
9	502.5	37.6	419	1 CBPA_BOVIN	P00730 bos taurus
10	488.5	36.5	419	1 CBPL_RAT	P00731 rattus norv
11	483	36.1	417	1 CBP2_RAT	P19222 rattus norv
12	465.5	34.8	419	1 CBP1_HUMAN	P15085 homo sapien
13	458	34.2	421	1 CBP4_HUMAN	P09142 homo sapien
14	456	34.1	417	1 CBP2_HUMAN	P48052 homo sapien
15	411.5	30.8	304	1 CBP2_STMYI	P42788 stimulum vi
16	381.5	28.5	433	1 CBPA_ANOGA	P02350 anopheles g
17	375.5	28.1	303	1 CBPB_ASFEI	P04069 astacus flu
18	286.5	21.4	424	1 CBPT_THEVU	P29068 thermacin
19	286.5	21.4	430	1 YHT2_YEAST	P18836 saccharomyc
20	282.5	21.1	451	1 CBPS_STRGR	P18143 streptomyc
21	255	19.1	434	1 CBPS_STRCP	P33041 streptomyc
22	103.5	7.7	444	1 TRSH_CHICK	P70080 gallus gail
23	99	7.4	444	1 TRSH_HUMAN	P17752 homo sapien
24	97	7.2	297	1 Y103_HUMAN	P15006 homo sapien
25	96.5	7.2	376	1 YOGT_BACSU	P54497 bacillus su
26	95	7.1	444	1 TRSH_RABIT	P17290 oryctolagus
27	95	7.1	707	1 ORCL_SCHPO	P54789 schizosacch
28	91	6.8	262	1 TCUL_ECOLI	P51983 escherichia
29	90.5	6.8	434	1 CBPH_BOVIN	P04836 bos taurus
30	90.5	6.8	476	1 CBPH_RAT	P15087 rattus norv
31	89.5	6.7	476	1 CBPH_MOUSE	P09242 xenopus lae
32	89	6.7	481	1 TRSH_MOUSE	P09242 xenopus lae
33	88.5	6.6	476	1 CBPH_HUMAN	P16870 homo sapien

34	86	6.4	382	1 HMBP_DROME	P22809 drosophila
35	86	6.4	447	1 TRSH_MOUSE	P17532 mus musculu
36	85.5	6.4	465	1 TRHA_ECOLI	P42787 drosophila
37	85.5	6.4	1406	1 CBPX_DROME	P15055 homo sapien
38	85	6.4	1255	1 PER2_HUMAN	P27206 bacillus su
39	84	6.3	3587	1 SRP1_BACSU	P32906 caenorhabdi
40	83.5	6.2	319	1 UN30_CAEEL	P09810 rattus norv
41	83.5	6.2	444	1 TRSH_RAT	P37892 lophius ame
42	83	6.2	454	1 CBPH_LOPAM	P38890 saccharomyc
43	81.5	6.1	526	1 YH07_YEAST	O42091 anguilla an
44	80.5	6.0	488	1 TY3H_ANGAN	P15169 homo sapien
45	80	6.0	458	1 CBPN_HUMAN	

ALIGNMENTS

RESULT	ID	CBPB_RAT	STANDARD:	PRT:	415 AA.
AC	CBPB_RAT	P19223;			
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-NOV-1990 (Rel. 16, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Carboxypeptidase B precursor (EC 3.4.17.2).				
GN	CPB.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=89034324; PubMed=3182872;				
RT	Clauser E., Gardell S.J., Craik G.S., Macdonald R.J., Rutter W.J.;				
RT	"Structural characterization of the rat carboxypeptidase A1 and B				
RT	genes. Comparative analysis of the rat carboxypeptidase gene				
RT	family."				
RL	J. Biol. Chem. 263:17837-17845(1988).				
CC	-I- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =				
CC	peptide + L-lysine(or L-arginine).				
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE				
CC	ZINC CARBOXYPEPTIDASE FAMILY.				
CC	-----				
CC	EMBL: M23959; AAA40872.1; JOINED.				
CC	EMBL: M23947; AAA40872.1; JOINED.				
CC	EMBL: M23950; AAA40872.1; JOINED.				
CC	EMBL: M23952; AAA40872.1; JOINED.				
CC	EMBL: M23953; AAA40872.1; JOINED.				
CC	EMBL: M23954; AAA40872.1; JOINED.				
CC	PIR: A32129; A32129.				
CC	HSSP: P09955; INSA.				
CC	MEMOPS: M14.003; -				
CC	InterPro: IPR003146; Propep_M14.				
CC	InterPro: IPR000834; Zn_Carboxypept.				
CC	Pfam: PF02244; Propep_M14; 1.				
CC	Pfam: PF00246; Zn_carboxypept. 1.				
CC	PRINTS: PR00765; CRBOXYPTASEA.				
CC	PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.				
CC	PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.				
CC	Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.				
CC	SIGNAL				
CC	PROPEP				
CC	CHAIN				
CC	METAL				
CC	METAL				

FT METAL 302 302 ZINC (BY SIMILARITY).
 FT ACT_SITE 376 376 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 171 184 BY SIMILARITY.
 FT DISULFID 243 266 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 SQ SEQUENCE 415 AA; 47515 MW; 8BA06CCADE30B6F2 CRC64;

Query Match 44.9%; Score 601; DB 1; Length 415;
 Best Local Similarity 44.6%; Pred. No. 6.8e-49;
 Matches 120; Conservative 42; Mismatches 67; Indels 40; Gaps 5;

QY 1 ASASYEQYSLNLTYSWIEFTTHERPDMLTIKIGSFEEKYPLVAVSGKEQTANNAI 60
 DB 109 ASGSHYTKNWTETEMIOQVATDNDPLVQSVIGTFEGRNMYLKI-GKTRPNKPAI 167
 QY 61 WIDGSHAREMISPAFLMFI-----GH----- 83
 DB 168 FIDCGFHAREMISPAFCOMFVREAVRTYNOEIHMKQLDELDFYVLPVNVINDGYVTWT 227
 QY 84 NRMWRKRSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVAS 143
 DB 228 DRMRKTRSTWAGSSCGVPRNRNF-NAGMCEVCASSPCSETYCGPAPSEKETKALAD 286
 QY 144 FLRNNIQIAVISMHSQHVPEYSTRSKSKDHELSLVAEVAIRKTSKNTRYT 203
 DB 287 FIRNNLSITIAVILHYSQMMLYPSYDYKLPENYELNALVGAARKEL-ATLHGKRYT 345
 QY 204 HGHSEETLYLAPGGDDMIYDLGIKYSFT 232
 DB 346 YGPGATTIYPAAGSDMSYDQIKIKYSFT 374

RESULT 2
 ID CBPB_BOVIN STANDARD; PRT; 306 AA.
 AC P00732;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase B (EC 3.4.17.2).
 GN CPB.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 ON NCBI_TaxID=9913;
 RX MEDLINE=75217824; PubMed=1057162;
 RA Tilani K., Ericsson L.H., Walsh K.A., Neurath H.;
 RT Amino-acid sequence of bovine carboxypeptidase B.";
 RL Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
 RN [2]
 RP SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
 RX MEDLINE=74260705; PubMed=4833744;
 RA Schmidt J.U., Hirs G.H.W.;
 RT Primary structure of bovine carboxypeptidase B. Inferences from the
 RT locations of the half-cystines and identification of the active site
 RT arginine.";
 RL J. Biol. Chem. 249:3756-3764(1974).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
 RX MEDLINE=76265065; PubMed=957425;
 RA Schmid M.F., Herriott J.R.;
 RT Structure of carboxypeptidase B at 2.8-A resolution.";
 RL J. Mol. Biol. 103:175-190(1976).
 RN [4]
 RP ACTIVE SITE.
 RX MEDLINE=70007159; PubMed=5344132;
 RA Plummer T.H., Jr.;
 RT Isolation and sequence of peptides at the active center of bovine
 RT carboxypeptidase B.";

RL J. Biol. Chem. 244:5246-5253(1969).
 RN [5]
 RP ACTIVE SITE.
 RX MEDLINE=73061487; PubMed=4565668;
 RA Kimmel M.T., Plummer T.H., Jr.;
 RT Identification of a glutamic acid at the active center of bovine
 RT carboxypeptidase B.";
 RL J. Biol. Chem. 247:7864-7869(1972).
 CC -1 CATALYTIC ACTIVITY: peptidyl-L-L-lysine(or L-arginine) + H(2)O =
 CC peptide + L-lysine(or L-arginine).
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC PIR; A00912; CPEOB.
 DR PDB; 1CPB; 30-SEP-83.
 DR MEROPS; M14.003; -.
 DR InterPro; IPR000834; Zn_carboxypept.
 DR Pfam; PF00246; Zn_carboxypept. 1.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
 FT DISULFID 63 76
 FT DISULFID 135 158
 FT METAL 149 163
 FT METAL 66 66
 FT METAL 69 69
 FT METAL 194 194
 FT ACT_SITE 246 246
 FT ACT_SITE 268 268
 SQ SEQUENCE 306 AA; 34612 MW; C329D2655C44A172 CRC64;

Query Match 43.8%; Score 586; DB 1; Length 306;
 Best Local Similarity 43.9%; Pred. No. 1.2e-47;
 Matches 116; Conservative 44; Mismatches 64; Indels 40; Gaps 5;

QY 6 YEQYHSNLTYSWIEFTTHERPDMLTIKIGSFEEKYPLVAVSGKEQTANNAIWDG 65
 DB 6 YEKNNMETEIAWTEQVASENPDLISRAIGTFIAGTYILLVY-GPGSNKRAVFDG 64
 QY 66 IHAREMISPAFLMFI-----GH-----NRMWR 88
 DB 65 FHAREMISPAFCOMFVREAVRTYNGREIHMTFLDKIDFYVLPVNVINDGYITWTNMR 124
 QY 89 KNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVASFLRN 148
 DB 125 KTRSTRAGSSCTGTDLNRNF-DAGWCISGASNNPCSETYCGSAASESEKSAVAADTRNH 183
 QY 149 INQIKAVISMHSQHVPEYSTRSKSKDHELSLVAEVAIRKTSKNTRYTHGHS 208
 DB 184 LSSIKAVLTHYSQMMLYPSYDYKLPNNVELNLAKGAVKKL-ASLHGTYISGPGA 242
 QY 209 EETLYLAPGGDDMIYDLGIKYSFT 232
 DB 243 TTYIPASGSDMAVDQIKIKYSFT 266

RESULT 3
 ID CBPB_CANFA STANDARD; PRT; 416 AA.
 AC P55261;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
 DE membrane associated protein) (ZAP47).
 GN CPB1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 ON NCBI_TaxID=9615;
 RX [1]
 RP SEQUENCE FROM N.A.

FT METAL 161 161 ZINC.
 FT METAL 164 164 ZINC.
 FT METAL 289 289 ZINC.
 FT ACT_SITE 341 341
 FT ACT_SITE 363 363 NUCLEOPHILE.
 FT STRAND 14 17
 FT HELIX 20 31
 FT TURN 32 32
 FT HELIX 43 45
 FT STRAND 50 52
 FT HELIX 61 69
 FT TURN 70 72
 FT STRAND 75 76
 SO SEQUENCE 401 AA; 45713 MW; 53129AFL159A26348 CRC64;

Query Match 43.0%; Score 575; DB 1; Length 401;
 Best Local Similarity 43.6%; Pred. No. 1.8e-46;
 Matches 115; Conservative 45; Mismatches 64; Indels 40; Gaps 5;

OY 6 YEYHSLNEIYSWIEFTEHRHDM.LTKIHGSSFEKYP.LYLVKSGKQOTAKNAIWDG 65
 Db 101 YEKYNNMETIEMTEQVYKSNPDLSRISGTFEGDNIVILKV-GKPGSKNPALFMDG 159
 OY 66 IHAREWISPARCLWFI-----GH-----NRMR 88
 Db 160 FHAREWISQACQWVFAVRYTEYEAHMEFELNDLDFYVLPVINDGVIYTWTKNRMR 219
 OY 89 KNRSPYANNHCIGDNLNRFASKHMGCEGSSSCSEYCGLYPSEPEYKAVASFLRRN 148
 Db 220 KTRSTNNGSSCTGTDPNRFN-NAAGMCTYGASVNPNCNETYGSAAESETETALDFIRNN 278
 OY 149 INOKATISMHSYQHYFEPYSTRSKSKDHELSLVASEAVRAIEKTSKNTRTYHGS 208
 Db 279 LSSIKAVLTHYSYQMLPYSDYKLPENDAEELNSLAKAVKEL-ASLYGTSYSGGS 337
 OY 209 ETLVAPGDDMIYDGIKYSFT 232
 Db 338 TTYTPAAGCSDDMAYNNGIKYSFT 361

RESULT 5
 CBP_HUMAN STANDARD; PRT; 417 AA.
 AC P15086; 060834;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein) (PASP).
 GN CPB1 OR CPB OR PCPB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
 RC TISSUE=Pancreas;
 RX MEDLINE=92129345; PubMed=1370825;
 RA Yamamoto K.K., Poussette A., Chow P., Wilson H., el Shami S., French C.K.;
 RA "Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of pancreas-specific protein as pancreatic procarboxypeptidase B.";
 RT J. Biol. Chem. 267:2575-2581(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=98182241; PubMed=9524066;
 RA Aloy P., Catusas L., Villegas V., Reverter D., Vendrell J., Aviles F.X.;
 RA "Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B.";

RL Biol. Chem. 379:149-155(1998).
 RN [3]
 RP SEQUENCE OF 16-43.
 RC TISSUE=Pancreas;
 RX MEDLINE=89153096; PubMed=2920728;
 RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
 RT "Purification and properties of five different forms of human procarboxypeptidases.";
 RL Eur. J. Biochem. 179:609-616(1989).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-lysine(or L-arginine) + H(2O) = peptide + L-lysine(or L-arginine).
 CC -1- TISSUE SPECIFICITY: PANCREAS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE ZINC CARBOXYPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M81057; AAA66973.1; -;
 DR EMBL; AJ224866; CAA12163.1; -;
 DR PIR; S02812; S02812.
 DR PIR; A42332; A42332.
 DR HSSP; P09955; INSA.
 DR MEROPS; M14.003; -;
 DR MTM; 114852; -;
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_Carboxypept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF02246; Zn_Carboxypept; 1.
 DR PRINTS; PR00765; CARBOXYPEPASE.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 DR KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL 1 15
 FT PROPEP 16 110
 FT CHAIN 111 417
 FT METAL 176 176
 FT METAL 179 179
 FT METAL 304 304
 FT ACT_SITE 378 378
 FT DISULFID 173 186
 FT DISULFID 245 268
 FT DISULFID 259 273
 FT CONFLICT 16 16
 FT CONFLICT 17 17
 FT CONFLICT 37 37
 FT CONFLICT 208 208
 FT CONFLICT 245 245
 SO SEQUENCE 417 AA; 47366 MW; BBLCF212D830305E CRC64;

Query Match 42.4%; Score 567; DB 1; Length 417;
 Best Local Similarity 41.3%; Pred. No. 1.1e-45;
 Matches 111; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

OY 1 ASASYEYHSLNEIYSWIEFTEHRHDM.LTKIHGSSFEKYP.LYLVKSGKQOTAKNAI 60
 Db 111 ATGSHTEKYNKMTETIEWTOQVATENDALISRSVIGTTEGRAIVILKV-GKAGQNKPAI 169
 OY 61 WIDGCIHAREWISPARCLWFI-----GH----- 83
 Db 170 FMDGCFHAREWISPARCLWFI-----GH----- 83
 OY 84 NNRMRKNSRYANNHCIGDNLNRFASKHMGCEGSSSCSEYCGLYPSEPEYKAVAS 143
 Db 230 SRFWKRTSRTHGSSCIGDPPNRF-DAQWCEIGASRNCDDEYCGPAESEKETALAD 288
 OY 144 FLRRNINQIKAVISMHSYQHYFEPYSTRSKSKDHELSLVASEAVRAIEKTSKNTRTY 203

```

DB 289 FIRMUSKATLTHSHSOMMITYSTAYKIGENNAELNALAKATVKEL-ASLHGCTKT 347
QY 204 HGHGSETLYLAPGGDDMYDGLGIRYSFT 232
DB 348 YGPATITVPAAGSGDDWADYDGGIRYSFT 376

RESULT 6
CBPC_MOUSE STANDARD: PRT: 417 AA.
ID CBPC_MOUSE
AC P15089:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90062123; PubMed=2584208;
RA Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA Seyaiz W.E.;
RT "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT novel member of the carboxypeptidase gene family."
RL J. Biol. Chem. 264:20094-20099(1989)
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: J05118; AAA37369.1; -.
DR PIR: A34487; A34487.
DR HSSP: P09955; INSA.
DR MEROPS: M14.010; -.
DR MGD: MGI:88479; CpB3.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carbopept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carbopept. 1.
DR PRINTS: PR00765; CARBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE.
FT CHAIN 110 417 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 245 268 BY SIMILARITY.
SQ SEQUENCE 417 AA; 48790 MW; A2B300A066BD1BA6D CRC64;

```

Query Match 41.9%; Score 561; DB 1; Length 417;
 Best Local Similarity 43.7%; Pred. NO. 3.9e-45;
 Matches 115; Conservative 37; Mismatches 71; Indels 40; Gaps 5;

```

QY 6 YEOYHSLNLTYSWIEFTRRHPDMLTKIHGSSFEYPLIVLKVSCKEOTAKNAIWDGC 65
DB 116 YAKYNDMDIVSWTEKMLHPEWRSIRKIGSTVEDNPPLYLKI-GKKDGERKAIPMDG 174
QY 66 IHAREWISPAFCIMFI-----GHN-----RMWR 88
DB 175 IHAREWISPAFCIMFIYQATKSGKKNKIMTKLLDRNPFVLPYFNVDGIWSTQDRMR 234
QY 89 KNRSEYANNHCIGTDLNRNPFASKHCEGASSSCSEYCGLYPESEPEVKAASFRLRN 148
DB 235 KNRSRNQNSTCIGTDLNRNF-DVSMDSPTNKNPCNLNVRGPAPESEKETKATNFIKSH 293
QY 149 INOKIYISMHSYSHIVPEYSTRKSKDHELSLVASEAVAIKTSKNRTYTHGHS 208
DB 294 LNSIKRYIFHSYSOMLLPYGTFFKLPNNHDLKVARIAVDAL-STRETRYIYGPIA 352
QY 209 EFLYLAPGGDDMYDGLGIRYSF 231
DB 353 STIYKTSGLDWYDGLGIRYF 375

RESULT 7
CBPC_RAT STANDARD: PRT: 309 AA.
ID CBPC_RAT
AC P21961:
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE A3).
GN CPA3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RX MEDLINE=91105153; PubMed=1988052;
RA Cole K.R., Kumar S., Le Trong H., Woodbury R.G., Walsh K.A.,
RA Neuraath H.;
RT "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT enzyme activity within mast cell granules."
RL Biochemistry 30:648-655(1991).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
DR PIR: A33118; A33118.
DR HSSP: P09955; INSA.
DR MEROPS: M14.010; -.
DR InterPro: IPR000834; Zn_carbopept.
DR Pfam: PF00246; Zn_carbopept. 1.
DR PRINTS: PR00765; CARBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; zinc.
FT METAL 68 68 ZINC (BY SIMILARITY).
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 196 196 ZINC (BY SIMILARITY).
FT ACT_SITE 248 248 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 270 270 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 65 78 BY SIMILARITY.
FT DISULFID 137 160 BY SIMILARITY.
FT VARIANT 1 1 MISSING (IN THE MAJOR FORM).
SQ SEQUENCE 309 AA; 35786 MW; 20330FABC3EB3EF CRC64;

```

Query Match 41.3%; Score 552; DB 1; Length 309;
 Best Local Similarity 42.6%; Pred. NO. 1.8e-44;
 Matches 112; Conservative 39; Mismatches 72; Indels 40; Gaps 5;

GN CPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91151335; PubMed=1998496;
RA le Huerou I., Guilloteau P., Touleec R., Pulgserver A., Wicker C.;
RT "Cloning and nucleotide sequence of a bovine pancreatic
RL preprocarboxypeptidase A cDNA.";
RN Biochem. Biophys. Res. Commun. 175:110-116(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=96096549; PubMed=8522204;
RA Goo J.H., Kim K.H., Choi K.Y.;
RT "Cloning, sequencing and expression of the gene encoding a major
RL alleotypic preprocarboxypeptidase A from bovine pancreas.";
RN Gene 165:333-334(1995).
RN [3]
RP SEQUENCE OF 111-417.
RX MEDLINE=71155183; PubMed=5102489;
RA Bradshaw R.A., Walsh K.A., Neurath H.;
RT "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
RN chymotryptic peptides of the cyanogen bromide fragment F.I.";
RL Biochemistry 10:938-950(1971).
RN [4]
RP REVISIONS TO 138 AND 141.
RX MEDLINE=72138789; PubMed=5143102;
RA Petra P.H., Hermanson M.A., Walsh K.A., Neurath H.;
RT "Characterization of bovine carboxypeptidase A (Allan).";
RN Biochemistry 10:4023-4025(1971).
RN [5]
RP SEQUENCE OF 17-120.
RX MEDLINE=89150306; PubMed=3147705;
RA Wade R.D., Haas G.M., Kumar S., Walsh K.A., Neurath H.;
RT "The amino acid sequence of the activation peptide of bovine pro-
RN carboxypeptidase A.";
RL Biochimie 70:1137-1142(1988).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
RX MEDLINE=83294519; PubMed=6887246;
RA Rees D.C., Lewis M., Lipscomb W.N.;
RT "Refined crystal structure of carboxypeptidase A at 1.54-A
RN resolution.";
RL J. Mol. Biol. 168:367-387(1983).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
RX MEDLINE=96003618; PubMed=7556081;
RA Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
RT "The three-dimensional structure of the native ternary complex of
RN bovine pancreatic procarboxypeptidase A with proproteinase E and
RL chymotrypsinogen C.";
RN EMBO J. 14:4387-4394(1995).
RN [8]
RP VARIANT ALLELIC.
RX MEDLINE=69283620; PubMed=5817619;
RA Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
RT "Identification of the amino acid replacements characterizing the
RN alleotypic forms of bovine carboxypeptidase A.";
RL Biochemistry 8:2762-2768(1969).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBUNIT: MONOMER. THE ZYMOGEN IS SECRETED AS A TERNARY COMPLEX
CC COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
CC PROTEINASE E.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -1- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/COA.html".
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M61851; AAA30426.1; -
DR EMBL; M61851; AAA30427.1; -
DR EMBL; Z33906; CAA83955.1; -
DR PIR; JN0126; CPBDA.
DR PIR; A31406; A31406.
DR PDB; 3CPA; 15-JAN-87.
DR PDB; 4CPA; 22-OCT-84.
DR PDB; 5CPA; 15-JAN-87.
DR PDB; 6CPA; 15-OCT-91.
DR PDB; 7CPA; 31-JAN-94.
DR PDB; 8CPA; 31-JAN-94.
DR PDB; 1CBX; 31-JAN-94.
DR PDB; 1CPS; 15-OCT-94.
DR PDB; 2CTB; 31-JAN-94.
DR PDB; 2CRC; 31-JAN-94.
DR PDB; 1ARL; 01-AUG-96.
DR PDB; 1ARM; 17-AUG-96.
DR PDB; 1BAV; 01-APR-97.
DR PDB; 1YME; 12-FEB-97.
DR PDB; 1CPX; 05-AUG-98.
DR PDB; 1PYT; 27-JAN-97.
DR MEROPS; M14.001; -
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept. 1.
DR PRINTS; PR00765; CROBOXPEPTASE.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal;
KW 3d-structure; Polymorphism.
FT SIGNAL 1 16
FT PROPEP 17 110
FT CHAIN 111 419
FT METAL 179 179
FT METAL 182 182
FT METAL 306 306
FT ACT_SITE 358 358
FT ACT_SITE 380 380
FT DISULFID 248 271
FT VARIANT 289 289
FT VARIANT 338 338
FT VARIANT 415 415
FT CONFLICT 95 95
FT CONFLICT 199 199
FT CONFLICT 203 203
FT CONFLICT 224 224
FT CONFLICT 232 232
FT CONFLICT 295 295
FT TURN 114 116
FT TURN 119 120
FT TURN 125 138
FT HELIX 140 142
FT STRAND 143 150
FT STRAND 152 153
FT STRAND 156 162
FT STRAND 171 176
FT TURN 180 181
FT HELIX 183 199
FT TURN 200 202
FT HELIX 204 212
FT STRAND 214 218
FT HELIX 223 231
FT TURN 232 232
FT TURN 234 235
I -> V (IN ALLELIC VARIANT).
E -> A (IN ALLELIC VARIANT).
L -> V (IN ALLELIC VARIANT).
S -> L (IN REF. 5).
D -> N (IN REF. 3).
D -> N (IN REF. 3).
D -> N (IN REF. 3).
O -> E (IN REF. 3).
D -> N (IN REF. 3).
ACTIVATION PEPTIDE.
CARBOXYPEPTIDASE A.
ZINC.
ZINC.
ZINC.
PROTON DONOR.
NUCLEOPHILE.

Query March	37.68;	Score 502.5;	DB 1;	Length 419;
Best Local Similarity	40.28;	Pred. No. 1.2e-39;		
Matches 106; Conservative	42;	Mismatches 75;	Indels 41;	Gaps 5

RESULT	10	CBP1_RAT	ID	STANDARD:	PRT;	419 AA.
CBP1_RAT			P00731			
DT	21-JUL-1986			(Rel. 01, Created)		
DT	01-FEB-1996			(Rel. 30, Last sequence update)		
DT	16-OCT-2001			(Rel. 40, Last annotation update)		
CD	Carboxypeptidase A1 precursor			(EC 3.4.17.1).		

CCOC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
NOX NCR1_TaxID=10116;
NRN [1]
RRP SEQUENCE FROM N.A.
RRP MEDLINE=82105986; PubMed=6275388;
RRA Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standering D.N.,
RRA Pitter R.L., Valenzuela P., Rutter W.J.,
RRT Rat preprocarboxypeptidase A: cDNA sequence and preliminary
characterization of the gene,
RRL Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
NRN [2]

RP	SEQUENCE FROM N.A.
RX	MEDLINE=69034324; PubMed=3182872;
RA	Clausen E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.,
RT	"Structural characterization of the rat carboxypeptidase A1 and B
RT	genes. Comparative analysis of the rat carboxypeptidase gene
RT	family.";
RL	J. Biol. Chem. 263:17837-17845(1988).
CC	-1 CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC	amino acid.
CC	-1 SUBUNIT: MONOMER.
CC	-1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC	ZINC CARBOXYPEPTIDASE FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; V01232; CAA2542.1; -
DR	EMBL; J00713; AAA40893.1; -
DR	EMBL; M23990; AAA40955.1; -
DR	EMBL; M23960; AAA40955.1; JOINED.
DR	EMBL; M23985; AAA40955.1; JOINED.
DR	EMBL; M23985; AAA40955.1; JOINED.
DR	EMBL; M23987; AAA40955.1; JOINED.
DR	EMBL; M23988; AAA40955.1; JOINED.
DR	EMBL; M23989; AAA40955.1; JOINED.
DR	PIR; A00911; CPPTA.
DR	PIR; B32129; B32129.
DR	HSP; P00730; IPYT.
DR	PIR; B32129; B32129.
DR	MEROPS: M14.001; -
DR	InterPro: IPR003146; Propep_M14.
DR	InterPro: IPR000834; Zn_carlopept.
DR	Pfam: PF02244; Propep_M14; 1.
DR	Pfam: PF02246; Zn_carlopept; 1.
DR	PRINTS: PR00765; CRBOXYPTASEA.
DR	PROSITE: PS00132; CARBOXYDEPT_ZN_1; 1.
DR	PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW	Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; signal.
FT	SIGNAL
FT	1
FT	PROPEP
FT	111 110
FT	CHAIN
FT	111 419
FT	METAL
FT	179 179
FT	METAL
FT	182 182
FT	METAL
FT	306 306
FT	ACT_SITE
FT	358 358
FT	ACT_SITE
FT	380 380
FT	DISULFID
FT	248 271
FT	CONFLICT
FT	196 196
FT	CONFLICT
FT	261 263
FT	CONFLICT
FT	347 347
SO	SEQUENCE
SO	419 AA; 47197 MW; BB002D1CB9B7491 CRC64;

[illegible]

OY 149 INQKAYISMHSYSHIVIPYSTRKSKDHELSIVASEAVAIKTSKNTRYTHGHS 208
 DB 297 GN-IFKAFISHSQLLLYGYTSEPAPQDELQDLAKSAVATL-TSLHGTFKFGYSII 354
 OY 209 EFLYLAPGGDDMIYDLGIRKYSFT 232
 DB 355 DRIYQASGSDIMWTYSGIKRYSFT 378

RESULT 11
 CBP2_RAT STANDARD; PRT; 417 AA.
 AC P19222;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
 GN CPA2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=pancreas;
 RX MEDLINE=89034323; PubMed=3182871;
 RA Gargiell S.J., Craik C.S., Clausner E., Goldsmith E.J., Stewart C.-B.,
 RA Graf M., Rutter W.J.;
 RT "A novel rat carboxypeptidase, CPA2: characterization, molecular
 RT cloning, and evolutionary implications on substrate specificity in
 RT the carboxypeptidase gene family.";
 RL J. Biol. Chem. 265:17828-17836(1988).
 RN [2]
 RP SEQUENCE OF 131-143 FROM N.A.
 RX MEDLINE=95386501; PubMed=7657630;
 RA Normant E., Gros C., Schwartz J.C.;
 RT "Carboxypeptidase A isoforms produced by distinct genes or
 RT alternative splicing in brain and other extrapancreatic tissues.";
 RL J. Biol. Chem. 270:20543-20549(1995).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RX MEDLINE=92105124; PubMed=1761558;
 RA Fleming Z., Kobe B., Stewart C.-B., Rutter W.J., Goldsmith E.J.;
 RT "Structural evolution of an enzyme specificity. The structure of rat
 RT carboxypeptidase A2 at 1.9-A resolution.";
 RL J. Biol. Chem. 266:24606-24612(1991).
 CC -1- CARLYTIC ACTIVITY: Similar to that of carboxypeptidase A
 CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
 CC residues.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M23721; AAA40956.1;
 DR EMBL; M23714; AAA40956.1; JOINED.
 DR EMBL; M23715; AAA40956.1; JOINED.
 DR EMBL; M23716; AAA40956.1; JOINED.
 DR EMBL; M23717; AAA40956.1; JOINED.
 DR EMBL; M23718; AAA40956.1; JOINED.
 DR EMBL; M23719; AAA40956.1; JOINED.
 DR EMBL; M23720; AAA40956.1; JOINED.
 DR EMBL; S79837; -; NOT_ANNOTATED_CDS.
 DR PIR; A32128; A32128.
 DR HSSP; P48052; IAYE.
 DR MEROPS; M14.002; -.
 DR InterPro; IPR003146; Proper_M14.

DR InterPro; IPR000834; Zn_Carboxypept.
 DR Pfam; PF02244; Proper_M14; 1.
 DR Pfam; PF02246; Zn_Carboxypept; 1.
 DR PRINTS; PR00765; CARBOXYPTASEA.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL; 1
 FT PROPEP; 17 112
 FT CHAIN; 113 417
 FT METAL; 177 177
 FT METAL; 180 180
 FT METAL; 304 304
 FT ACT_SITE; 378 378
 FT DISULFID; 246 269
 FT DISULFID; 318 352
 SO SEQUENCE 417 AA; 46912 MW; 180041A830F2D45 CRC64;

Query Match 36.1%; Score 483; DB 1; Length 417;
 Best Local Similarity 38.2%; Pred. No. 8e-38;
 Matches 109; Conservative 47; Mismatches 75; Indels 54; Gaps 9;

OY 6 YEOYHSLNETIYSMTPEETTERHPMLTKIHIGSSPEKPLVLAWS-GKEQTAKNAIWDIC 64
 DB 118 FEAYHTLEETIYQEDNDLVANPGLVSKVNLGSSFENRPMNVLFSTGDD--KPAIWLDA 174
 OY 65 GHAHREWISPAFCILM-----FT-----GH-----NRMW 87
 DB 175 GHAREWVOTATMTANKIASDYGTPATLTLNTIDIFLPTVNDGIVFSQTTNRMW 234
 OY 88 RKNRSFYANNHCIGTDLNRNFASKHWCESGASSSCSEYCGLYPESEPEVKAVASFLRR 147
 DB 235 RKRFRSKRSGSCGVDPENRNM-DANFGPGASSPSCSDSYHGPKPNEVEVKSYIDPIKS 293
 OY 148 NINQKAYISMHSYSHIVIPYSTRKSKDHELSIVASEAVAIKTSKNTRYTHGHS 207
 DB 294 H-GVKKAFITLHYSQLLMPYGYKCRKPPDFELDEVAQKAQAL-KRLHGTSYKVGPI 351
 OY 208 SETLYLAPGGDDMIYDLGIRKYSFTSN-----PYEKLTP 242
 DB 352 CSVIYQASGSDIMWYDLGIRKYSFTAFELRLRTAFYGLLPKQILP 396

RESULT 12
 CBP1_HUMAN STANDARD; PRT; 419 AA.
 AC P15085;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
 GN CPA1 OR CPA.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=pancreas;
 RX MEDLINE=93038569; PubMed=1417781;
 RA Catasus L., Villegas V., Pascual R., Aviles F.X.,
 RA Wacker-Planquart C., Pulgover A.;
 RT "cDNA cloning and sequence analysis of human pancreatic
 RT procarboxypeptidase A1.";
 RL Biochem. J. 287:299-303(1992).
 RN [2]
 RP SEQUENCE OF 17-42.
 RX MEDLINE=89153096; PubMed=2920728;
 RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
 RT "Purification and properties of five different forms of human
 RT procarboxypeptidases.";
 RL Eur. J. Biochem. 179:609-616(1989).


```

CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X67318; CAA47732.1; -.
DR PIR; S02810; S02810.
DR PIR; S29127; S29127.
DR HSSP; P00730; 1PYT.
DR MEROPS; M14.001; -.
DR MIM; 114850; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxpept; 1.
DR PRINTS; PR00765; CROBOXPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR HydroLase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1 16
FT PROPEP 17 110
FT CHAIN 111 419
FT METAL 179 179
FT METAL 182 182
FT METAL 306 306
FT ACT_SITE 358 358
FT ACT_SITE 380 380
FT DISULFID 248 271
FT SEQUENCE 419 AA; 47140 MW; 439FAFFAE958B1 CRC64;

Query Match 34.8%; Score 465.5; DB 1; Length 419;
Best Local Similarity 37.9%; Pred. No. 3.5e-36;
Matches 100; Conservative 44; Mismatches 79; Indels 41; Gaps 5;

D6 YEOHSLNEIYSWIEFTEHPDMLTKIHGSSFEKYLVLKSGKQTKAKNAIWDG 65
D6 119 YATHTLEETIDYDELVAENPHLVSKIQIGTYEGRIYLFKS-TGSGSKRAPIAIDTG 177
D6 66 IHAREWISPAFLCMTI-----GHNRMWR 88
D6 178 IHSREWTOAGGVFAKKIINDYGDAAFTAILDTLDFLEIVNPDGFATHTSTNRMR 237
D6 89 KNRSEYANNHCIGTDLRNENFASKHMCBEGASSSCSEYTCGLYPESEBEVAKAVASFLRN 148
D6 238 KTRSHYTGSLCIGVDPRNMDAGFGL-SGASSNPSCSEYTHGKFPANSEVEKSIYDVYDH 296
D6 149 INOKAVISMHSYSOHIVPEPYSTRSKSDHELSIVASEAVRAIEKTSKTRTHGHS 208
D6 297 GN-ITKAISHISQSLMAYPRGYTEPEVDDDELQLSKAVALT-ASLVGCTKKNYSII 354
D6 209 ETLVLAPGGDDMIYDLGIKYSFT 232
D6 355 KAIYQASGSTIDWYSGIKYSFT 378

RESULT 13
CBP4_HUMAN STANDARD; PRT; 421 AA.
AC G0U142;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).

```

```

GN CPB4 OR CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99310537; PubMed=10383164;
RA Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.I.;
RT "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT decarboxylase inhibitors during differentiation of prostate epithelial
RT cancer cells."
RL Cancer Res. 59:2981-2988(1999).
CC -1- FUNCTION: Could be involved in the histone hyperacetylation
CC pathway.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF095719; AAF23230.1; -.
DR HSSP; P48052; IAYE.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxpept; 1.
DR PRINTS; PR00765; CROBOXPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR HydroLase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1 16
FT PROPEP 17 113
FT CHAIN 114 421
FT METAL 181 181
FT METAL 184 184
FT METAL 308 308
FT ACT_SITE 382 382
FT ACT_SITE 382 382
FT DISULFID 250 273
FT SEQUENCE 421 AA; 47379 MW; 91883D5B705BC6A CRC64;

Query Match 34.2%; Score 458; DB 1; Length 421;
Best Local Similarity 38.4%; Pred. No. 1.8e-35;
Matches 103; Conservative 35; Mismatches 82; Indels 48; Gaps 6;

D6 YEOHSLNEIYSWIEFTEHPDMLTKIHGSSFEKYLVLKSGKQTKAKNAIWDG 65
D6 120 YGAYHSLSEIYHEMDNTLAADPDLARVYKIGHSPENRPMVLYKSTGKYRRAPVWLNAG 179
D6 66 IHAREWISPAFLCMTI-----FT-----GHNRMWR 88
D6 180 IHSREWISOATVAITWARKIYSDYQRPATISILEKMDIFILPVANPDGYVYTGQONLMR 239
D6 89 KNRSEYANNHCIGTDLRN-----FASKHMCBEGASSSCSEYTCGLYPESEBEVAKAVASFL 144
D6 240 KTRSRNPGSSCIGADPPRRNNASPAK-----GASDPCEVYHGPRANSEVEKSVYDF 294
D6 145 LRRNINQIKAYISMHSYSOHIVPEPYSTRSKSDHELSIVASEAVRAIEKTSKTRTHGHS 204
D6 295 IOKHGN-FKGFIDHSYSQSLMAYPRGYSVKKAQDAEELDKVAVLAKALASYS-GIEIYQV 352
D6 205 GHGSETIYLPAGGGDDMIYDLGIKYSFT 232
D6 353 GPCTTVYPAAGSSIDWAYDNGIKFAFT 380

RESULT 14

```


CPB2_HUMAN
ID CPB2_HUMAN STANDARD; PRT; 417 AA.
AC P48032;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN CPA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.
RC TISSUE=Pancreas;
RA MEDLINE=95204457; PubMed=7896805;
RA Catusus L., Vendrell J., Aviles F.X., Carreira S., Pulgarer A.,
RA Billeter M.;
RT "The sequence and conformation of human pancreatic
RT procarboxypeptidase A2. cDNA cloning, sequence analysis, and
RT three-dimensional model.";
RL J. Biol. Chem. 270:6651-6657(1995).
RN [2]
RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98111000; PubMed=9450539;
RX Reverter D., Garcia-Saez I., Catusus L., Vendrell J., Coll M.,
RX Aviles F.X.;
RT "Characterisation and preliminary X-ray diffraction analysis of human
RT pancreatic procarboxypeptidase A2.";
RL FEBS Lett. 420:7-10(1997).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98046021; PubMed=9384570;
RX Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT "The three-dimensional structure of human procarboxypeptidase A2.
RT Deciphering the basis of the inhibition, activation and intrinsic
RT activity of the zymogen.";
RL EMBO J. 16:6906-6913(1997).
CC -1- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC residues.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U19977; AAA74425.1; -
CC PDB: 1AYE; 13-JAN-99.
DR MEROPS: M14.002; -
DR MIM: 600688; -
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_Carboxpept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_Carboxpept. 1.
DR PRINTS: PR00765; CROXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW 3D-structure.
FT SIGNAL 1 16 POTENTIAL.
FT PROPEP 17 112 ACTIVATION PEPTIDE.
FT CHAIN 113 417 CARBOXYPEPTIDASE A2.
FT METAL 177 177 ZINC.
FT METAL 180 180 ZINC.
FT METAL 304 304 ZINC.
FT ACT_SITE 378 378 NUCLEOPHILE.
FT DISULFID 246 269

FT DISULFID 318 352
SQ SEQUENCE 417 AA; 46840 MW; 6D8107330A2DBC58 CRC64;
Query Match 34.1%; Score 456; DB 1; Length 417;
Best local Similarity 36.5%; Pred. No. 2.7e-35;
Matches 103; Conservative 45; Mismatches 80; Indels 54; Gaps 7;
QY 9 YHSLNEIYSWIEFITERHPDMLTKRHIGSSFEYRPLVLKVS-GKEQTAKNAIWDGIR 67
DB 121 YHRLTEEISQEMDLVAHEHPLGVSKVNISSFEHRPNMVLKFSIGD--KPAIWLDAIGH 177
QY 68 ARWISPAFLMPTG-----HNRMRKN 90
DB 178 ARWVTOATLMTANKIVSDYGRDPSITSLDALDIFLPTNPDPGVFSQTKRNRMRKT 237
QY 91 RSEFYANNHCITGLDNRNFASKHWCCEGASSSCSSEYCGILPSEPEYKVASFLRNIN 150
DB 238 RSKVSGSLCVDPNRRW-DAGFGCPASSNPDSYHGSANSSEVNSIVDFIKSH-G 295
QY 151 QIKAVISMHSQSIHVEPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYHGGSET 210
DB 296 KVKAFILHSQDLMPFYCYKTKTDPELSEVNAKAQSL-RSLHGKIKVGVPCSV 354
QY 211 LYLAPGGDDWIDYLGKISFTSN-----PYEKILP 242
DB 355 IYQASGSDIMSYDYGKYSFAFELRDTGRYGFLLPARQILP 396
RESULT 15
ID CPB2_SIMV1 STANDARD; PRT; 304 AA.
AC P42788;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc carboxypeptidase (EC 3.4.17.-) (fragment).
OS Simulium vittatum (Black fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Simuliidae; Simulium.
OX NCBI_TaxID=7192;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Gut;
RC MEDLINE=94093864; PubMed=8269093;
RX Ramos A., Mahowald A., Jacobs-Lorena M.;
RT "Gut-specific genes from the black fly *Simulium vittatum* encoding
RT trypsin-like and carboxypeptidase-like proteins.";
RT Insect Mol. Biol. 1:149-163(1993).
CC -1- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
CC -1- TISSUE SPECIFICITY: GUT-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: L08481; AAL8531.1; -
CC HSP: P48052; 1AYE.
DR InterPro: IPR000834; Zn_Carboxpept.
DR Pfam: PF00246; Zn_Carboxpept. 1.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc.
FT METAL 1 1 ZINC (BY SIMILARITY).
FT METAL 58 58 ZINC (BY SIMILARITY).
FT METAL 61 61 ZINC (BY SIMILARITY).

OM of: US-09-980-881-4 to: N_Geneseq_032802:* out_format : pfs
Date: Sep 18, 2002 7:03 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+ .p2n.model -DEV=x1h
-O=/cgn2.1/USPRO.spool/US09980881/runat_16092002_140130_10714/app-query.fasta_1.1126
-DB=N_Geneseq_032802 -OPMT=fastp -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOAPEXT=0.500
-GAPOP=4.500 -GAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum2
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09980881.ecgn1_1_0 -NCPU=6 -ICPU=3 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-4
Query length: 246
Database: N_Geneseq_032802:*
Database sequences: 1738436
Database length: 858457221
Search time (sec): 386.310000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC1962	1325.00	2699.78	5.0e-142	15	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV14302	1272.00	2592.93	4.4e-136	12	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV11671	1272.00	2589.27	7.1e-136	17	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2846	1272.00	2589.27	7.1e-136	17	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ11001	1267.00	2578.96	2.7e-135	17	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAT5760	610.00	1231.43	3.0e-60	92	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ30601	582.00	1173.77	5.0e-57	92	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ30600	582.00	1170.58	7.4e-57	12	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42500	567.00	1141.90	2.9e-55	99	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT4249	567.00	1141.30	3.2e-55	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT4294	567.00	1139.21	4.2e-55	12	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT4294	567.00	1139.21	4.2e-55	12	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42506	567.00	1139.02	4.3e-55	12	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAAT4804	560.00	1126.86	2.0e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42511	560.00	1126.80	2.0e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT4278	559.00	1124.73	2.7e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42512	559.00	1124.73	2.7e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAD03837	556.50	1117.13	7.1e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62806	556.00	1118.55	5.9e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62790	554.00	1114.42	1.0e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62801	553.00	1114.42	1.0e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62789	553.00	1112.36	1.3e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62791	552.00	1110.30	1.7e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62805	551.00	1108.24	2.2e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62803	551.00	1108.24	2.2e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62804	551.00	1108.24	2.2e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT62807	550.50	1106.51	2.8e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT6479	550.50	1106.51	2.8e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT6476	550.50	1104.48	3.6e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT6478	550.50	1104.48	3.6e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT61728	550.50	1100.94	5.6e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62808	549.00	1104.11	3.8e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62800	549.00	1104.11	3.8e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62800	549.00	1097.58	8.7e-53	18	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAAT62800	549.00	1097.58	8.7e-53	18	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62802	548.00	1092.95	4.9e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62802	548.00	1092.95	4.9e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62807	546.00	1097.93	8.3e-53	10	

/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAD03837	525.50	1047.63	5.3e-50	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAD03477	492.00	985.13	1.6e-46	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAD03838	471.50	944.40	3.0e-44	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC98830	465.50	929.98	1.9e-43	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAAT46119	465.50	929.95	1.9e-43	

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC1962

seq_documentation_block:

ID AAC81962 standard; cDNA; 1573 BP.
XX
XX AAC81962;
AC
XX
XX 01-MAR-2001 (first entry)
DT
XX
XX Human brain carboxypeptidase B cDNA.
DE
XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimer's; neurotrophic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 18..1100
FT
FT CDS /tag= a
FT /product= "carboxypeptidase B"
XX
XX PN WO20006717-A1.
XX
XX PD 09-NOV-2000.
XX
XX PF 01-MAY-2000; 2000WO-JP02878.
XX
XX PR 30-APR-1999; 99JP-0125169.
XX
XX (MATS/) MATSUMOTO A.
XX
XX Matsumoto A;
XX
XX WPI: 2000-687534/67.
XX
XX P-PSDB; AAB11457.
XX
XX Human brain carboxypeptidase B isolated from the hippocampus useful for
XX screening agents for the treatment of Alzheimer's and other brain
XX disorders -
XX
XX Claim 2b; Page 64-68; 84pp; Japanese.
PS
XX
XX
XX

This invention describes a novel protein with peptidase activity
against brain beta-amyloid precursor protein which has been isolated from
human hippocampus and which has cerebroprotective, antialzheimer's,
neurotrophic, neuroprotective and hemostatic activity and which can be used
as a vaccine or for gene therapy. The protein, and compounds identified
by screening as promoters or inhibitors of its activity, are used to
regulate beta-amyloid accumulation in the brain and treat or prevent
diseases in which this occurs, such as Alzheimer's, senile dementia,
inherited cerebral hemorrhage, Down's syndrome, and head trauma.

Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;

alignment_scores:

Quality:	Length:	Gaps:
1325.00	246	0
Ratio: 5.408		
Percent Similarity: 99.593	Percent Identity: 98.780	

alignment_block:

US-09-980-881-4 x AAC81962 ..
Align seg 1/1 to: AAC81962 from: 1 to: 1573

```

1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
  |||
360 GCCTCCGACATCGTACTATGAAACAGATACACTACTAATAGAAATCATATTC 409
17 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
  |||
410 TTGGATAGAAATTATTAAGTGAAGCATCTGATATGCTTACAAAAATCC 459
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
  |||
460 ACATTGGATCCCTCATTTTGAAGTAGTACCCCTCTATGTTTAAAGGTTTCT 509
51 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 67
  |||
510 GGAAGAAAGCAACAGCCAAAAATGCCATATGATGATGATGATGATGATCA 559
67 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisA 84
  |||
560 TGCCAGAGATGATGATCTCTCTGCTTCTGCTTGTGTTGATGATGATGATG 609
84 snArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
  |||
610 ATGCAATGTGGAGAAAGAACGCTTCTTCTATGCGAACATCATATGTCATC 659
101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluG 117
  |||
660 GGAACAGACCTGCAATAGCAACCTTCTCTCCAAACACGCTGTGAGAGAAAG 709
117 YAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGlu 134
  |||
710 TGCAATCCAGTCTCTATCTCGGAAACCTACTGCTGACTTATCTCTGAGT 759
134 ergLupProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 150
  |||
760 CAGAACCCAGAAAGTGAAGCAGTGGCTAGTTCTTGAGAAAGAAATCTAAC 809
151 GlnIleLysAlaIleTyrIleSerMetHisSerTyrSerGlnHisIleVal 167
  |||
810 CAGATTAAGCATACATCAGCATGCTATCATACATCCAGCATATATGCTGT 859
167 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 184
  |||
860 TCCATATTCCTATACAGCAAGTAAAGCAAGACCATAGAGACGCTCTC 909
184 euValAlaSerGluAlaValAlaArgAlaIleGluLysThrSerLysAsn 200
  |||
910 TAGTACCCAGTGAAGCAGTCTGCTATGACAAAACTAGTAAAAATATCC 959
201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyG 217
  |||
960 AGGTATACACATGCGCATGCTCAGAAACCTTATACCTAGCTCCGAGAG 1009
217 yGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThrSer 234
  |||
1010 TGGGAGAGATGATCTATGATTTGGGATCAAAATATTCGTTTACATCA 1059
234 snProProValGluLysLeuLeuProLeuSerLeuLys 246
  |||
1060 ACCCACCCTGTAGAGAAAGCTTTGGCGCTGTCTTAAAA 1097
seq_name: /stdsl/gcgdata/geneseq/geneseq-emb1/NM1999.DAT:AAV74302
seq_documentation_block:
ID AAV74302 standard; cDNA; 1272 BP.
XX
XX AAV74302;
AC
XX
XX 28-APR-1999 (first entry)
DE Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
XX
XX Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW

```

```

KW polymorphism detection; thrombotic disease; ds.
XX Homo sapiens.
OS
XX W09855645-A1.
PN
XX 10-DEC-1998.
PD
XX 02-JUN-1998; 98WO-EP03244.
PE
XX 03-JUN-1997; 97US-0869057.
PR
XX (SCHD ) SCHERING AG.
PA
XX Morser MJ, Nagashima M;
PI
XX WPI; 1999-045800/04.
DR
XX P-PSDB; AAW92270.
XX
XX Detecting new polymorphism of human plasma carboxypeptidase B -
PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
XX Example 1; Page 24; 35pp; English.
XX
XX This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphs of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphs in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphs within the general
CC population compared with populations known to be at risk and establishing
CC a PCPB polymorph profile for at-risk patients.
XX
XX Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;
SO

```

alignment_scores:

Quality:	1272.00	Length:	301
Ratio:	5.171	Gaps:	3
Percent Similarity:	81.728	Percent Identity:	81.395

alignment_block:

```

US-09-980-881-4 x AAV74302

```

Align seg 1/1 to: AAV74302 from: 1 to: 1272

```

1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
  |||
343 GCCTCCGACATCGTACTATGAAACAGATACACTACTAATAGAAATCATATTC 392
17 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
  |||
393 TTGGATAGAAATTATTAAGTGAAGCATCTGATATGCTTACAAAAATCC 442
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
  |||
443 ACATTGGATCCCTCATTTTGAAGTAGTACCCCTCTATGTTTAAAGGTTTCT 492
51 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 67
  |||
493 GGAAGAAAGCAACAGCCAAAAATGCCATATGATGATGATGATGATGATCA 542
67 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 83
  |||
543 TGCCAGAGATGATGATCTCTCTGCTTCTGCTTGTGTTGATGATGATGATG 592

```

```

83 ..... 83
593 TAACCAATTCATAGGGAATAGGCAATATACCAATCTCTAGAGCTT 642
83 ..... 83
643 GTGGATTTCATGTATWCCGGGTGTTAATGTGGACGGTTATGACTACTC 652
84 .....AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsn 97
693 ATGGAAAAAGAAATCGAATGTGGAGAAAGAACGCTTCTTCTATACGACACA 742
97 snHiscysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTSP 113
743 ATCATGTGATCGGACAGACCTGAAATAGGAATCTTCTTCCAAACACTGG 792
114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrGlyGlyLe 130
793 TGTGAGGAAGGTGCATCCAGTTCTCATGCTCGAAACCTACTGTGACT 842
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
843 TTATCTGTAGTCAGACACAGAGTGAAGGAGTGCTAGTTCTTGAGAA 892
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
893 GAAATATCAACAGATTAAAGCATATCATCAGCATTCATCATCTCCAC 942
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 180
943 CATATAGCTTTCCATATTCCTATACACAGATAAAGCAAAACACATGA 992
180 uGluSerLeuValAlaSerGluAlaValArgAlaIleGlyLysThr 197
993 GGACAGCTCTCTAGTAGCCAGTAGACAGTCTGCTATGAAAAACTA 1042
197 erLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
1043 GTAAAAATACAGAGTATACACATGCGCATGCGCAAAACCTTATACCTA 1092
214 AlaProGlyGlyLysAspTrpIleTyrAspLeuGlyLleLysTyrSe 230
1093 GCTCCTGGAGGTGGGACATGTGATGTGATTTGGCATCAAAATATTC 1142
230 r.Phe..... 231
1143 GTTACAATTCGACTCGAGATACGGGACATACGATTTCTGTGCGCG 1192
232 .....ThrSerAsnProProValGluLysLeuLeuProLeuSerLeu 246
1193 AGCGTTACATCAAAACCCACTGTAGAGAAAGCTTTTGCCTGTCTCTAAA 1242
246 s 246
1243 A 1243
seq_name: /SIDS1/gcgsdata/geneseq/geneseqn-emb1/NA1996.DAT:AA111671
seq_documentation_block:
ID AA111671 standard; DNA; 1749 BP.
XX
AC AA111671;
XX
DT 12-APR-1996 (first entry)
XX
DE Human plasma carboxypeptidase B coding sequence.
XX
KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers

```

```

FT CDS 41..1312
FT FT /*tag= a
FT FT /product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT FT /*tag= b
FT mat_peptide 107..1309
FT FT /*tag= c
XX
XX US5474901-A.
XX
XX 12-DEC-1995.
XX
XX 01-FEB-1991; 91US-0649591.
XX PF
XX 01-FEB-1991; 91US-0649591.
XX PR
XX 14-OCT-1992; 92US-0959944.
XX PR
XX 15-DEC-1993; 93US-0167727.
XX PR
XX 19-JUL-1994; 94US-0277540.
XX
XX (GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI; 1996-039508/04.
XX DR
XX P-PSDB; AAR90293.
XX
XX Antibody to human plasma carboxypeptidase B - useful for detecting
XX PT and purifying hPCPB for use in treating clotting disorders e.g.
XX PT haemophilia A
XX
XX PS Disclosure; Figure 4; 40pp; English.
XX
XX An antibody which specifically binds human plasma carboxypeptidase B
XX CC (hPCPB) and does not cross react with other carboxypeptidases is
XX CC useful for the detection of hPCPB in vitro. The antibody is also
XX CC used for purifying hPCPB from a sample. Purification comprises
XX CC passing a sample thought to contain hPCPB over either a column to
XX CC which antibody has been bound, or a plasminogen affinity column,
XX CC eluting the column and then recovering the fraction containing the
XX CC hPCPB.
XX
XX SQ Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

alignment_scores:
Quality: 1272.00 Length: 301
Ratio: 5.171 Gaps: 3
Percent Similarity: 81.728 Percent Identity: 81.395

alignment_block:
US-09-980-881-4 x AA111671 ..
Align seg 1/1 to: AA111671 from: 1 to: 1749

1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
383 GCCGCCGCATCGTACTATGAAAGATATCACTCACTAATGAAATCTATTTC 432
17 rTTPILleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
433 TTGGATAGAAATTTAACTAGAGAGCATCCGATATGCTTACAAAAATTC 482
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValIleLysValSer 50
483 ACATGATCTCTCATTTGAGAAAGTACCCACTCTATGTTTAAAGTTTCT 532
51 GlyLysGluGlnThrAlaLysAsnAlaIleTTPILeAspCysGlyIleH 67
533 GGAAGAAACAAACAGCCAAATGATGATGATGATGATGATGATGATGATG 582
67 sAlaArgGluTTPILeSerProAlaPheCysLeuTTPILeGlyHis. 83
583 TGCAGAGAAATGATCTCTCTGCTTTCTGTGTGTGTGTGTGTGTGTGTGT 632

```

```

83 ..... 83
633 TAACATATCTATGGGATATAGGCAATATACATCTCTGAGGCTT 682
83 ..... 83
683 GTGGATTCTATGTTATVCCGGTGAATGAGGAGGTATGACTACTC 732
84 .....AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnA 97
733 ATGAAAAAAGATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGA 782
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHis 113
783 ATCATTTGCAATCGGACAGACCTGAATAGAGACTTCTCTCAACAC 832
114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 130
833 TGTGAGAGAGGTGCATCCAGTCTCTCATGCTCGAAACCTACTGT 882
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeu 147
883 TTATCTGTGATCAGAACGAGAGTGAAGCAGTGGCTACTTCTTGAG 932
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSer 163
933 GAAATATCAACGATTAAGCATATCAGCATCATTCATCATCCAG 982
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAsn 180
983 CATATAGGTTTCCATATTCCTATACAGAGTAAAGCAAGCAACCA 1032
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLys 197
1033 GGAAGTGTCTAGTAGCCAGTAGAAGCAGTGTCTATTGGAAGACT 1082
197 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyr 213
1083 GTAAATAATACCGGTATACATGCGCATGGCTCAGAAACCTTAAC 1132
214 AlaProGlyGlyLysAspAspTrpIleTyrAspLeuGlyIleLys 230
1133 GCTCTGGAGGTGGGAGCATGTGATCATGATTGGGCAATCAATAT 1182
230 r.Phe..... 231
1183 GTTACAAATGAACTTCAGATACGGGCACATACGATTCCTGTGCC 1232
232 .....ThisSerAsnProProValGluLysLeuProLeuSerLeu 246
1233 ACCGTTACATCAAAACCCACTGTAGAGAAAGCTTTGGCCGCTGT 1282
246 s 246
1283 A 1283
seq_name: /STDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:AA62846
seq_documentation_block:
ID AA62846 standard; DNA: 1749 BP.
XX
AC AA62846;
XX
DT 08-MAY-1997 (first entry)
XX
DE Human plasma carboxypeptidase B coding sequence.
XX
KW Human: plasma carboxypeptidase B; PCPB; haemostatic regulation;
XX plasma; plasminogen; ss.
XX
OS Homo sapiens.
XX

```

```

FH Key Location/Qualifiers
FT CDS 41..1312
FT ../*tag= a
FT .. /product= Human PCPB
FT sig_peptide 41..106
FT ../*tag= b
FT mat_peptide 107..1309
FT ../*tag= c
FT misc_binding 134..177
FT ../*tag= d
FT .. /bound_moiety= 46_bp_probe
XX
XX US593674-A.
XX
XX 14-JAN-1997.
XX
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 14-OCT-1992; 92US-0959944.
XX
XX 15-DEC-1993; 93US-0167727.
XX
XX 19-JUL-1994; 94US-0277540.
XX
XX 27-APR-1995; 95US-0430787.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI: 1997-099413/09.
XX
XX P-PSDB: AAM14733.
XX
XX Using human plasma carboxypeptidase B in blood coagulation - is
XX functionally related to carboxypeptidase A and pancreas
XX carboxypeptidase B
XX
XX Example 2: Column 37-42; 39pp; English.
XX
XX This sequence encodes human plasma carboxypeptidase B (PCPB) which
XX has a molecular weight under non-reducing SDS-PAGE of approx. 60 kd.
XX PCPB may be used therapeutically in haemostatic regulation. PCPB is
XX purified from human plasma or by transformed cell culture by
XX extraction using plasminogen bound to a solid phase.
XX
XX Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;
XX
XX
XX alignment_scores:
XX Quality: 1272.00 Length: 301
XX Ratio: 5.171 Gaps: 3
XX Percent Similarity: 81.728 Percent Identity: 81.395
XX
XX alignment_block:
XX US-09-980-881-4 x AA62846 ..
XX
XX Align seg 1/1 to: AA62846 from: 1 to: 1749
XX
XX 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
XX |||||
XX 383 GCCTCCGCAATGCTACTATGAAACAGTATCAGTCAATGAAATCTATTC 432
XX |||||
XX 17 rTrpIleGluPheIleThrGlnAlaGlnHisProAspMetLeuThrLysIleH 34
XX |||||
XX 433 TTGGATGAAATTTATTAAGTGAAGAGCATCCGATATGCTTACAAATCC 482
XX |||||
XX 34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
XX |||||
XX 483 ACATTGATCTCTCATTTGAGAGAGTACCACCTCTATGTTTAAAGTTTCT 532
XX |||||
XX 51 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 67
XX |||||
XX 533 GGAAGAGAACAAACAGCCAAATGATGATGATGATGATGATGATGATGAT 582
XX |||||
XX 67 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 83

```



```

|||||
583 TCCAGAGATGATCTCTCTGCTTGTGCTGTTGATAGGCCATA 632
83 ..... 83
633 TAACTCATTTATGGGATATAGGCAATATACCAATCTCCGAGGCTT 682
83 ..... 83
683 GTGGATTTCTATGATCCGGTGTAAATGAGACGTTATGACTACTC 732
84 .....AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnA 97
733 ATGAAAAAGATCGAATCTGGAGAAAGAACCGTTCTTCTATGCGAACA 782
97 snHsCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHs1TP 113
783 ATCAATGTCATCGGAACAGACCTGAAATAGAACTTGCTTCCAAACACTG 832
114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
833 TGTGAGGAAGGTGCATGCAGTTCCTCATGCTCGAAACCTACTGTGACT 882
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
883 TTATCTGTAGTCAGAACCAAGAGTGAAGCACTGGCTACTTCTTGAGAA 932
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
933 GAAATATCAACCAAGATTAAGCATATCAGCATCATTCATATCCACG 982
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHs1G 180
983 CATATAGCTTTCCATATTCCTATACAGAACTAAAGCAAGACCATCA 1032
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197
1033 GGAACCTGCTCTAGTAGGACAGTGAAGCACTTGTCTATTTGAAACTA 1082
197 erLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
1083 GTAAAAAACAGAGTATACACATGCGCATGGCTCAGAAACCTTAATACCTA 1132
214 AlaProGlyGlyLysAspAspTrpIleTyrAspLeuGlyLysTyrSe 230
1133 GCTCTGACAGCTGGGACAGATTGATCTATGATTTGGCATCAATATATTC 1182
230 r.Phe..... 231
1183 GTTTCACATTTGAATCTGAGATACGGGACATACGATTTCTGTGCGCG 1232
232 .....ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuL 246
1233 AGCGTTACATCAAAACCCACTGTAGAGAACTTTTCCGCGTGTCTTAAA 1282
246 s 246
1283 A 1283
seq.name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAF18005
seq_documentation_block:
ID AAF18005 standard; DNA; 1400 BP.
XX
AC AAF18005;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 24.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
cardioactive; immunomodulatory; muscular active; vulnery;
gastrointestinal; nephrotropic; antinfective; gynecological;

```

```

KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
PN WO200055180-A2.
XX
21-SEP-2000.
XX
08-MAR-2000; 2000WO-US05918.
XX
12-MAR-1999; 99US-0124270.
XX
(PA) (HUMA-) HUMAN GENOME SCI INC.
XX
(ROSE/) ROSEN C A.
XX
PI Ruben SM:
XX
DR WP1: 2000-587514/55.
DR P-PSDB: AAB58129.
XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer.
XX
PS Claim 1; Page 507; 1425pp; English.
XX
CC Polynucleotide sequences AAF1982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences. Their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnery; gastrointestinal
CC general; nephrotropic; antinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;
XX
alignment_scores:
Quality: 902.00 Length: 238
Ratio: 4.983 Gaps: 4
Percent Similarity: 76.050 Percent Identity: 75.210
alignment_block:
US-09-980-881-4 x AAF18005 ..
Align seg 1/1 to: AAF18005 from: 1 to: 1400
65 GYLIIENISAlaArgLutRPIleSerProAlaPheCysLeuTrpPheI 81
|||||
10 GGAATCCATGCCAGANNAATGATCTCTGCTTCTTGTGTTGATCAT 59
|||||
81 eGLYHs..... 83
60 AGGCCATATATCAATTCATGGGATAATAGGCAATATACCAATCTCC 109
|||||
83 ..... 83
110 TGAGCTTGCGATTTCTATGCTTATGCCGCTGTTAATGCGATGCTTAT 159
84 .....Asn.ArgMetTrpArgLysAsnArgSerPheT 94
::: |||||||||||||||||||||||||||

```



```

711 TGAGTGAATGCGCTGCTAAGACTGTAAGAAACTT...GCCCTAC 757
197 gxlxasnrhrarqlytrhrhlglyhlglysergluThrleuTyrlu 213
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
758 TGCAGCGGACCAAGTACATATGCGCCGCGAGCTACACATCTTCTCT 807
214 AlAProGlyGlyGlyAspAspTrpIleTyAspLeuGlyTleTyTyrSe 230
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
808 GCGCTGCGGGGCTGTGACGACTGCGCTTATGACCAAGAAATCATATTC 857
230 rPheThr 232
|||||
858 CTTCACC 864

seq_name: /SID51/gcdata/geneseq/geneseq-emb1/NA1996.DAT:AA142494
seq_documentation_block:
ID AA142494 standard; DNA; 1263 BP.
XX
XX AA142494;
XX
XX 12-FEB-1997 (first entry)
XX
XX Human pancreatic carboxypeptidase B encoding sequence.
DE
XX
XX ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW mustard-ribonuclease; antibody directed enzyme prodng therapy;
KW anti-neoplastic; prodng; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW F(ab')2; Pe1B; leader; human carboxypeptidase B; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX CDS 1..1248
XX FT /*tag= a
XX FT /product= preproenzyme
XX FT 1..39
XX FT /*tag= b
XX FT /note= "partial enzyme secretion leader sequence
XX FT (pre-sequence); residues -108 to -96"
XX
XX sig_peptide 40..324
XX FT /*tag= c
XX FT /note= "pro-sequence; residues -95 to -1"
XX FT 385..1245
XX FT /*tag= d
XX FT /note= "mature enzyme; encodes residues +1 to +307"
XX
XX MO9620011-A1.
XX
XX PD 04-JUL-1996.
XX
XX PE 21-DEC-1995; 95WO-G802991.
XX
XX 16-AUG-1995; 95GB-0016810.
XX PR 23-DEC-1994; 94GB-0026192.
XX
XX (ZENE ) ZENeca LTD.
XX
XX Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
XX Hennam JF, Hennequin LFA, Marsham PR, Rablin BR, Slater AM;
XX Farragone-Fiol A, Taylorson CJ;
XX
XX WPI; 1996-321650/32.
XX DR P-PSDB; AAW06172.
XX
XX Two component system for anti-tumour therapy - comprising targeting
XX PT moiety linked to mutated enzyme which can transform an
XX PT anti-neoplastic prodng
XX
XX Reference Example 14; Page 131; 182pp; English.
XX
XX A two-component system for anti-tumour therapy comprises a targeting
XX
XX

```

```

CC moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC prodng. The system is based on antibody directed enzyme prodng therapy
CC (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC pref. human pancreatic ribonuclease (HP-RNase), (see AA142478-83).
CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC be used. The present sequence encodes the native prepro-HCPB (see
CC AAW06172) contained in vector pIC11698.
XX
XX Sequence 1263 BP; 343 A; 306 C; 302 G; 312 T; 0 other;
SQ

alignment_scores:
Quality: 567.00 Length: 269
Ratio: 3.133 Gaps: 5
Percent Similarity: 67.286 Percent Identity: 41.264

alignment_block:
US-09-980-881-4 x AA142494 ..
Align seg 1/1 to: AA142494 from: 1 to: 1263

1 AlAserAlaSerTyTrpGluGlnTrpHisSerLeuAsnGluIleTyrSe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
325 GCAACAGGACACAGATTATGAGAAATGACAAAGTGGAAACGATGAGAGC 374
17 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
375 TTGGACTCAACAAGTCGCACACGAGATCCACCTCATCTTCCGACAG 424
34 IsIleGlySerSerPheGluLysTyProLeuTyValLeuLysValSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
425 TTATCGAGACACATTTGAGGACGCGCTATTTACTCTCGTAAGGTT... 471
51 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspGlyIleH 67
||||| ||| |||||:|||||:|||||:|||||:|||||:|||||
472 GGCAGAGCTGGACAAATTAAGCTGCTCATTTCAATGAGCTGGTTTCCA 521
67 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle..... 81
||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
522 TCCAGAGAGTGAGATTTCCTCGCATTCGCCAGIGGTTTGAAGAGAG 571
81 ..... 81
572 CTGTTGTAACATGAGAGTGATCCAAAGTGCAGAGCTTTCGACAAAG 621
82 ..... 83
622 TTAGACTTTATGTCCTGCTGCTCAATATTGATGCTACATCTACAC 671
84 .....AsnArgMetTrpArgLysAsnArgSerPheTyraIaAsnA 97
672 CTGACCAAGAGCCGATTTTGGAGAAAGACGCGCCACCCATCTGGAT 721
97 snHiscysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrp 113
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||
722 TAGCTGATTTGGCAGACAGCCCAAGAAATTTT...GATCGTGGTTGG 768
114 CysGluGluGlyAlaSerSerSerCysSerGluTrpTrpCysGlyLe 130
||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
769 TGTGAATTTGAGAGCTCTCGAAACCCCTGTGATGAATTTCTGTGGACC 818
130 uTyProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
819 TGCAGCAGAGTCGAAAGAGACCAAGGCGCTGCGATTTCAATCCGCA 868
147 rGAsnIleAsnGluIleLysAlaTyrlleSerMetHisSerTySerGln 163
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||
869 ACAAACTCTTCCATCAAGGATATGACAAATCCACTGCTACTCCCAA 918
164 HisIleValPheProTySerTyThrArgSerLysSerLysAspHisG 180
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||
919 ATGATGATCTACCTTACTCATATGCTTACAAACCTGCTGAGAACATGC 968

```

```

180 uGIuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197
|||||.....:|||||.....:|||||.....:
969 TGAGTTGAATGCGCTGAAGCTGTAAGAAAGACTT...GCCATC 1015
197 erlyasnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
:::|||||.....:|||||.....:|||||.....:
1016 TGCACGGCAGCAAGTACATATGCGCCGGAGCTACACATATCTATCT 1065
214 AlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSe 230
||| |||||.....:|||||.....:|||||.....:
1066 GCTGCTGGGGCTCTGACGACTGCGCTTATGACCAAGCAATCATATTC 1115
230 rPheThr 232
|||||
1116 CTTTCACC 1122

seq_name: /sids1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV41795
seq_documentation_block:
ID AAV41795 standard; DNA; 1263 BP.
XX AAV41795;
AC AAV41795;
DF 20-NOV-1998 (first entry)
XX 20-NOV-1998 (first entry)
DE Human pancreatic carboxypeptidase nucleotide sequence.
XX ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KM produg therapy.
XX Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..1248
FT /tag= a
FT /product= "Pancreatic carboxypeptidase B"
FT /note= "No start codon given"
XX W09835988-A1.
XX 20-AUG-1998.
XX 10-FEB-1998; 98WO-GB00415.
XX 29-OCT-1997; 97GB-0022727.
XX 14-FEB-1997; 97GB-0003104.
XX 18-OCT-1997; 97GB-0022003.
XX (ZENE ) ZENEPA (ZENE ) ZENECA LTD.
XX Edge MD;
XX WPI: 1998-467168/40.
XX P-PSDB; AAW74476.
XX New modified pro-domain of carboxypeptidase B - enhances expression
XX of co-expressed proteins for production of recombinant
XX carboxypeptidase or its fusions with antibodies, used, e.g. in
XX enzyme produg therapy
XX Example 1; Page 52-53; 83pp; English.
XX The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
XX modified pro-domain of CPB on a separate gene to enhance recombinant
XX expression. This process can be used to produce recombinant CPB in
XX eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
XX insulin production and protein sequencing, while its fusions with
XX antibodies are useful in antibody-directed enzyme produg therapy. The
XX modified pro-domain provide increased yields of recombinant CPB, possibly
XX by protecting the C-terminus against enzymatic degradation or by
XX increasing intracellular trafficking.
XX Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;

```

```

alignment_scores:
  Quality: 567.00      Length: 269
  Ratio: 3.133
  Percent Similarity: 67.286      Percent Identity: 41.264

alignment_block:
us-09-980-881-4 x AAV41795 ..
Align seg 1/1 to: AAV41795 from: 1 to: 1263

1 AlaserAlaserTyrGluGluThrHisSerLeuAsnGluLysTyrSe 17
|||||.....:|||||.....:|||||.....:
325 GCAACAGGACACAGTTATGAGAGATCAACAAGTGGAAAGCAATAGAGCC 374
17 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrIleH 34
||| |||.....:|||||.....:|||||.....:
375 TTGACTCATACAGTCCGCTGAGAAATCCAGCCCTCATCTCGCAGTG 424
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
|||||.....:|||||.....:|||||.....:
425 TTATCGAACCACATTTGAGGAGCGGCTATTACCTCGTGAAGTT... 471
51 GlyLysGluGluThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 67
||||| ||| |||||.....:|||||.....:|||||.....:
472 GGCAAAGCTGACCAAAATTAAGCTGCCATTTCATGAGCTGCTTCCA 521
67 sAlaArgIleTrpIleSerProAlaPheCysLeuTrpPheIle..... 81
|||||.....:|||||.....:|||||.....:
522 TGCCAGAGATGATTTCTCGCATTCGCCAGTGTTGTAGAGAGG 571
81 .....:|||||.....:|||||.....:
572 CTGTTCGACCTATGAGAGTGAGATCCAAAGTACAGAGCTTCTGACAAAG 621
82 .....:|||||.....:|||||.....:
622 TTAGACTTTTATGCTCTGCTGCTCAATATGATGCTCATCTACATC 671
84 .....:AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnA 97
672 CTGACCAAGAGCCGATTTTGGAGAGAGCTGCTCCACCAATCTGAT 721
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrp 113
::: |||||.....:|||||.....:|||||.....:
722 CTAGCTGATTTGGCAGACAGCCCAAGAAATTTT...GATGCTGTTGG 768
114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
769 TGTGAATTTGGAGCCTTCGAAGAACCCCTGTGATGAACCTTACTG 818
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
819 TGCAGCAGAGTCTGAAAGAAACCAAGCCCTGCTATTTCATTCGCA 868
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
|||||.....:|||||.....:|||||.....:|||||.....:
869 ACAAACTCTCTTCATCAAGCATATCTGACATCATCTGATCCCA 918
164 HisIleValPheProTyrSerTyrThrArgSerLysLysAspHisG1 180
|||||.....:|||||.....:|||||.....:
919 ATGATGATCTACCCCTTACTCATATGCTTACAACACCGTGACAAAC 968
180 uGluSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197
|||||.....:|||||.....:|||||.....:
969 TGAGTTGAATGCGCTGAAGCTGTAAGAAAGACTT...GCCATC 1015
197 erlyasnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
|||||.....:|||||.....:|||||.....:
1016 TGCACGGCAGCAAGTACATATGCGCCGGAGCTACACATATCTATCT 1065
214 AlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSe 230

```

```

|||||
1066 GCTGCTGGGGCTCTGACGACTGGCTTATGACCAAGAAATCAGATTATTC 1115
230 rphenr 232
1116 CTCACCC 1122

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AA142506
seq_documentation_block:
ID AA142506 standard; DNA: 1284 BP.
AC AA142506;
XX
XX 17-FEB-1997 (first entry)
XX
XX PROHCPB gene with PelB leader sequence.
XX
XX ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
XX mustard-ribonucleotide; antibody directed enzyme prodng therapy;
XX anti-neoplastic; prodng; reverse polarity; ion pair interaction;
XX reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
XX F(ab')2; PelB; leader; human carboxypeptidase B; ss.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..1275
XX FT /*tag- a
XX FT sig_peptide 1..66
XX FT /*tag- b
XX FT /*note- "PelB leader sequence"
XX FT sig_peptide 67..351
XX FT /*tag- c
XX FT /*note- "pro sequence"
XX FT mat_peptide 352..1272
XX FT /*tag- d
XX FT /*note- "mature HCPB sequence"
XX
XX MO9620011-A1.
XX
XX 04-JUL-1996.
XX
XX 21-DEC-1995; 95WO-GB02991.
XX
XX 16-AUG-1995; 95GB-0016810.
XX 23-DEC-1994; 94GB-0026192.
XX
XX (ZENE ) ZENECA LTD.
XX
XX Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
XX Hennam JF, Hennequin LFA, Marsham PR, Radin BR, Slater AM;
XX Tarraagon-Fiol A, Taylorson CJ;
XX
XX WPI; 1996-321650/32.
XX P-PSDB; AAW06175.
XX
XX Two component system for anti-tumour therapy - comprising targeting
XX moiety linked to mutated enzyme which can transform an
XX anti-neoplastic prodng
XX
XX Reference Example 18; Page 140-142; 182pp; English.
XX
XX A two-component system for anti-tumour therapy comprises a targeting
XX moiety linked to a mutated enzyme which can transform an anti-neoplastic
XX prodng. The system is based on antibody directed enzyme prodng therapy
XX (ADEPT) using non-naturally occurring mutant forms of host enzymes,
XX pref. human pancreatic ribonuclease (HP-RNase), (see AA142478-83).
XX Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
XX be used. The present sequence is a cloned pro-HCPB gene contained in
XX plasmid pIC11738 and which can be expressed in E. coli.
XX
XX Sequence 1284 BP; 355 A; 315 C; 299 G; 315 T; 0 other;

```

```

alignment_scores:
Quality: 567.00 Length: 269
Ratio: 3.133 Gaps: 5
Percent Similarity: 67.286 Percent Identity: 41.264

alignment_block:
US-09-980-881-4 x AA142506 ..
Align seg 1/1 to: AA142506 from: 1 to: 1284

1 AlaSerAlaSerTyrTyrGluGluTyrHisSerLeuAsnGluLeuTyrSe 17
|||||
352 GCACAGAGACACAGATTGAGAAAGTACAAAGTGGAAACGATACAGGC 401
17 rTTPILGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
|||||
402 TTGGACTCAACAGATGCGCAGTGAATCCAGCCCATCTCTCGCAGTG 451
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
|||||
452 TATCGAACAACATTGAGGAGCGGCTATTACTCTGTAAGTT... 498
51 GlyLysGluGluThrAlaLysAsnAlaIleTyrPheAspCysGlyLeuH 67
|||||
499 GCAAAAGCTGACAAATAAGCTGCCATTTTCATGACTGTGTTTCA 548
67 sAlaArgGluTyrPheSerProAlaPheCysLeuTyrPheIle..... 81
549 TGCCAGAGAGTGATTTCTGCAATTCGCCAGGTGTTGTAAGAGAGC 598
81 .....
599 CMTTGTGACGTATGAGCTGAGATCCAAAGTCCAGAGCTTTCAGACAG 648
82 .....GlyHis..... 83
649 TTAGACTTTATGCTCGCTGCTGCATATTTGATGGCTATACATAC 698
84 .....AsnArgMetTyrPheLysAsnArgSerPheTyrAlaAsnA 97
699 CTGGACCAAGACCGATTTTGGAAAGACCTGCTCCACCCATCTGAT 748
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTyr 113
:: |||||
749 CTAGCTGATTTGGCAAGACCCCAACAGAAATTT...GATCTGTGTG 795
114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
|||||
796 TGTGAATTTGGAGCTCTGAAACCCTGTGATGAACCTTACTGTGAGC 845
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
846 TGCCCCAGAGCTTGAAAGAGACCAAGCCCTGCTATTTTCATCCGA 895
147 rGAsnIleAsnGluIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
|||||
896 ACAAACTCTCTTCATCAAGGATATCTGACATTCATCTGACTGCCAA 945
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 180
|||||
946 ATGATGATCTACCCCTTACTCATATGCTTCAAAACCTGGAGAACATGC 995
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrS 197
|||||
996 TGAGTTGAAATGCCCTGCTAAAGCTACTGTGAAGAATCT...GCCATAC 1042
197 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
|||||
1043 TGACAGCGACCAAGTACATATAGCCCGGAGAGCTACAAACAATATACCT 1092
214 AlaProGlyGlyGlyAspAspTyrPheTyrAspLeuGlyIleLysTyrSe 230

```


This Page Blank (uspto)

OM of: US-09-980-881-4 to: Issued_Patents_NA.* out_format : pfs

Date: Sep 18, 2002 6:56 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet_p2n.model -DEV=x1h
-O=/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-920-812-13 + 81.50 120.09 53.18
-DB=Issued_Patents_NA -OPMT=fastap -SUFPR=ini -GAPOP=12.000
-GAPEXT=4.000 -MIMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blousum62
-TRANS=Numan40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEADSIZE=500 -MILEN=0 -MAXLEN=2000000000
-USER=US09980881 -ECGN1_1_62 -NCPU=6 -ICPU=3 -LONGLOS
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPR3 -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-4
Query length: 246
Database: Issued_Patents_NA.*
Database sequences: 383533
Database length: 122816752
Search time (sec): 90.520000

score_list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation	..	1.8e-147	1272	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-869-057-1 +		1272.00	2785.71	1.8e-147	1272	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-07-649-591B-2 +		1272.00	2781.78	2.9e-147	1749	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-277-540-2 +		1272.00	2781.78	2.9e-147	1749	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-430-787A-2 +		1272.00	2781.78	2.9e-147	1749	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-782-760-5 +		610.00	1321.16	6.7e-66	927	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-696-139-3 +		582.00	1259.13	1.9e-62	921	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-696-139-1 +		582.00	1255.71	3.0e-62	1215	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-67 +		567.00	1224.86	1.5e-60	989	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-64 +		567.00	1224.21	1.7e-60	1053	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-56 +		567.00	1221.76	2.2e-60	1263	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-71 +		567.00	1221.76	2.3e-60	1284	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-74 +		567.00	1208.61	1.7e-59	1059	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-77 +		559.00	1206.39	1.7e-59	1059	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-171-945-112 +		549.00	1177.19	8.8e-58	1870	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-171-945-124 +		549.00	1175.44	8.8e-58	2154	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-640-906-1 +		465.50	996.87	7.7e-48	1257	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-395-936-1 +		465.50	996.87	7.7e-48	1257	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-640-906-3 +		459.00	982.51	4.9e-47	1251	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-395-936-3 +		459.00	982.51	4.9e-47	1251	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-998-416-125 -		169.50	351.30	7.0e-12	515	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-331-709-2 +		163.00	335.18	5.6e-11	591	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-706-216-1 +		96.50	168.83	0.1026	2719	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-484-105-15 +		95.00	166.52	0.1379	2504	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-484-106-15 +		95.00	166.52	0.1379	2504	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-689-974-2 -		88.50	157.82	0.4211	1576	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-058-376-2 +		88.50	157.82	0.4211	1576	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-233-989-1 +		88.50	152.42	0.8408	2439	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-453-262-1 +		88.50	152.40	0.8429	2443	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-734-550-1 +		88.50	152.40	0.8429	2443	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-073-297-1 +		88.50	152.40	0.8429	2443	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-073-297-1 +		88.00	158.78	0.3724	1333	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-609-049A-27 +		86.50	135.27	7.59	6831	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-170-996-27 +		86.50	135.27	7.59	6831	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-966-363-10 -		85.00	144.26	2.40	2520	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-966-363A-10 -		85.00	144.26	2.40	2520	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-297-928-6 -		85.00	144.26	2.40	2520	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-07-903-710-1 +		82.50	157.80	0.4219	537	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-252-966B-1 +		82.50	157.80	0.4219	537	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-222-638B-1 +		82.50	157.80	0.4219	537	1			

/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-920-812-13 + 81.50 120.09 53.18
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-920-827-13 + 81.50 120.09 53.18
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-921-177-13 + 81.50 120.09 53.18
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-362-577C-13 + 81.50 120.09 53.18
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-920-828-13 + 81.50 120.09 53.18

seq_name: /cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-869-057-1

seq_documentation_block:

Sequence 1, Application US/08869057
Patent No. 5985562
GENERAL INFORMATION:
APPLICANT: Moser, Michael J
TITLE OF INVENTION: Method of Detecting Thrombotic Disease
TITLE OF INVENTION: Risk
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Berlex Biosciences Legal Department
STREET: 15049 San Pablo Avenue
CITY: Richmond
STATE: California
COUNTRY: USA
ZIP: 94804-0099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,057
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Washien, Wendy L.
REGISTRATION NUMBER: 36,301
REFERENCE/DOCKET NUMBER: 51509AUSM1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-262-5411
TELEFAX: 510-262-7095
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1272 base pairs
TYPE: nucleic acid
TOPOLOGY: double
MOLTYPE: linear
MOLWGT: 1272
PUBLICATION INFORMATION:
AUTHORS: Eaton, Dan L.
AUTHORS: Malloy, Beth E.
AUTHORS: Tsai, Siao P
AUTHORS: Henzel, William
AUTHORS: Dryden, Dennis
TITLE: Isolation, Molecular Cloning, and Partial
TITLE: Characterization of a No. 5985562el Carboxypeptidase B
TITLE: from Human Plasma
JOURNAL: J. Biol. Chem.
VOLUME: 266
ISSUE: 32
PAGES: 21833-21838
DATE: No. 5985562 15-1991
US-08-869-057-1
alignment_scores:
Quality: 1272.00 Length: 301
Ratio: 5.171 Gaps: 3
Percent Similarity: 81.728 Percent Identity: 81.395
alignment_block:
US-09-980-881-4 x US-08-869-057-1 ..
Align seg 1/1 to: US-08-869-057-1 from: 1 to: 1272

```

1  AAserAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
   |||||||
343 GCGTCGCGCATCGTACTATGTAACAGTATCACTCACTAAATGAATCTATTC 392
17  rTTPileGluPheIleThrGluArgHisProAspMetLeuThrIleHis 34
   |||||||
393 TTGGATAGAAATTTATTAAGTACGAGGCGATCTCTGATATGCTTACAAATATCC 442
34  lAlIleGlySerSerPheGluTyrTyrProLeuTyrValLeuLysValSer 50
   |||||||
443 ACATTGATCTCATTTGAGAAAGTACCACTCATGTTTAAAGGTTCT 492
51  GlYlyGluGlnThrAlaLysAsnAlaIleTPIleAspCysGlyIleH 67
   |||||||
493 GGAAGAAAGCAACAGCAAAATGCAATGATGATGACTGTGGAATCCA 542
67  sAlaArgGluTPIleSerProAlaPheCysLeuTyrPheIleGlyHis 83
   |||||||
543 TCCCAAGAAATGATCTCTGCTTCTGCTTGTGTTGTTCAATAGGCCATA 592
83  .....
593 TAACTCAATTTCTATGGATATATAGGCAATATACCAATCTCGTAGGCTT 642
83  .....
643 GTGATTTCTATGTTATGCCGGTGTAAATGTGAGCGGTTAGACTATCTC 692
84  .....AsnArgMetTrrPArgLysAsnArgSerPheTyrAlaAsn 97
   |||||||
693 ATGAAAAAAGATCGAATGTGAGAAAGACGTTCTTCTATGCGAACA 742
97  sNHlscYsAlIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisItrp 113
   |||||||
743 ATCATTTGCATTCGAAACAGACCTGAATAGCAACTTCTTCCAACACATGG 792
114 CysGluGluGlnAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
   |||||||
793 TGTGAGGAAGTGTCATCCAGTTCTCTCATGCTGGAAACCTACTGTGGACT 842
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
   |||||||
843 TTATCTGATCAGCAACAGCAAGAGTGAAGCACTGCTGTTGAGAGA 892
147 rGAsnIleAsnGlnIleLysAlaIleTyrIleSerMetHisSerTyrSerGln 163
   |||||||
893 GAAATATCAACCGATTAAGCATATCATCAGCATTCATTCATACCTCCAG 942
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG1 180
   |||||||
943 CATATAGTGTTCATATTCCTATATACAGAGTAAAGCAAGCAACCATGA 992
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrS 197
   |||||||
993 GGAAGTGTCTAGTACGAGGAGGAGTGTGCTCTATTTGAGAAACTA 1042
197 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
   |||||||
1043 GTAAAAATACAGAGGTATACATGCGCTGCTGAGAAACCTTATATACCTA 1092
214 AlaProGluGlyLysAspTrrPleTyrAspLeuGlyIleLysTyrSe 230
   |||||||
1093 GCTCTGAGAGTGGGAGCATTTGATCTATGATTTGGCATCAATATATTC 1142
230 r.Phe.....
   |||
1143 GTTAAACAATTCAGATTCAGATACGGGCAATACGAGATTTCTGGTGGCGG 1192
232 .....ThiSerAsnProProValGluLysLeuLeuProLeuSerLeuL 246
   |||||||
1193 AGCGTTACATCAAAACCCACCTGTAGAGAAAGCTTTGGCGCTGTCTCTAAA 1242

```

246 s 246
1
1243 A 1243

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-649-591B-2

```

seq_documentation_block:
: Sequence 2, Application US/07649591B
: Patent No. 5206161
: GENERAL INFORMATION:
: APPLICANT: Dennis Drayna and Daniel Eaton
: TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/649,591B
: FILING DATE: 19910201
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 689
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/266-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1749 bases
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: hybridization probe
: LOCATION: 133 to 178
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: FEATURE:
: NAME/KEY: potential clip site
: LOCATION: 380 to 382
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: FEATURE:
: NAME/KEY: signal sequence
: LOCATION: 41 to 106
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: US-07-649-591B-2

```

alignment_scores:

Quality	Ratio:	Length:
1272.00	5.171	301
Percent Similarity: 81.728		Gaps: 3
Percent Identity: 81.395		

alignment_block:

US-09-980-881-4 x US-07-649-591B-2 ..

Align seg 1/1 to: US-07-649-591B-2 from: 1 to: 1749

```

1 AlaserAlaserTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
  |||
383 GCCTCCGCAATCGACTATGAAACAGTATCACTCAATAATGAATCTATATC 432
17 rTrrPllGluPheIleThrGluArgHisProAspMetLeuThrIleH 34
  |||
433 TTGGATGAAATTTTAACTGAGAGGCACTCTGATATGCTTACAAAATGCC 482
34 IsIleGlySerSerPheGluIleTyrProLeuTyrValLeuIleValSer 50
  |||
483 ACATTGGATTCCTCATTTGAGAACTACCACTCATGTTTAAAGGTTCT 532
51 GlyIleGluGlnThrAlaIysAsnAlaIleTrrIleAspCysGlyIleH 67
  |||
533 GGAAGAAACAAACAGCAACCAAAATGCCATGATGCTGAGTGAATCCA 582
67 sAlaArgGluTrrIleSerProAlaPheCysLeuTrrPheIleGlyHis 83
  |||
583 TGCAGAGAAATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 632
83 .....
633 TAACCAATTCTATGGGATATAGGCAATATACCAATCTCTGAGGCTT 682
83 .....
683 GTCGATTTCTATGTATGCCGCTGTTAATGTGACGGTTATGACTACTC 732
84 ..... AsnArgMetTrrPArgIysAsnArgSerPheTyrAlaAsn 97
  |||
733 ATGGAAGAAACAAATCGAATGTGAGAAAGAACCTTCTTCTATGCGAACA 782
97 snHisCysIleGlyTrrAspLeuAsnArgAsnPheAlaSerLysHisTrr 113
  |||
783 ATCATTTGCATCGGACAGACACTGAATAGAACTTGTCTTCCAAACACTGG 832
114 CysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLe 130
  |||
833 TGTGAGAGAGGTGCATCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 882
130 uTyrProGluSerGluProGluValIleValAlaIleAlaSerPheLeuArg 147
  |||
883 TTATTCCTGATCGAAGCAAGCAAGTGAAGGCAAGTGTCTTCTTGAGAA 932
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
  |||
933 GAAATATCAACCGATTAAGCATACATCGATGATTCATATCTCCAG 982
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 180
  |||
983 CATATAGTGTTCATTCATTCCTATACAGCAAGTAAAGCAAGACCATGA 1032
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197
  |||
1033 GGAAGCTGTCTAGTACCGAGTGAAGCAGTGTCTGCTATTGAGAAAACCTA 1082
197 eTrrLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
  |||
1083 GTAAATAATCCAGTATACATGCGCATGCGTCAGAAACCTTATACCTA 1132
214 AlaProGlyGlyIleAspAspTrrPleTrrAspLeuGlyIleLysTyrSe 230
  |||
1133 GCCTCGAGAGGTGGGACGATGATCTATGATTTGGCAATCAAAATATTC 1182
230 r.Phe.....
1183 GTTTACAAATGAACTCGAGATACGGGCACATACGATTTCTGCTGCCGG 1232
232 .....ThSerAsnProProValGluIleLeuLeuProLeuSerLeu 246
  |||
1233 AGGGTATCAATCAACCAACCTCTGTAGAGAACCTTTTGGCGGTCTGTCTAAA 1282

```

```

246 s 246
1283 A 1283

seq_name: /cgn2_6/pdata/1/ina/5A_COMB.seq:US-08-277-540-2
seq_documentation_block:
; Sequence 2, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: NO. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: patin (genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-277-540-2

alignment_scores:
Quality: 1272.00 Length: 301
Ratio: 5.171 Gaps: 3
Percent Similarity: 81.728 Percent Identity: 81.395

alignment_block:
US-09-980-881-4 x US-08-277-540-2 ..
Align seg 1/1 to: US-08-277-540-2 from: 1 to: 1749

1 AlaserAlaserTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
  |||
383 GCCTCCGCAATCGACTATGAAACAGTATCACTCAATAATGAATCTATATC 432
17 rTrrPllGluPheIleThrGluArgHisProAspMetLeuThrIleH 34
  |||
433 TTGGATGAAATTTTAACTGAGAGGCACTCTGATATGCTTACAAAATGCC 482

```

```

34 1sileglserserphgclulysrtyrproleutyralleulysvalser 50
|||||
483 ACATTTGGATCTCATTTTGAGAACTACCCACTATATGTTTAAAGGTTTCT 532
51 G1LYSG1UG1nThrAlaLysAsnAla1leTrrPleAspCysGlyIleH1 67
|||||
533 GGAAGAAAGACAAACAGCCAAAATGCCATATGGATTGACGTGGAAATCCA 582
67 sAlaArGgluTrPleSerProAlaPheCysLeuTrPheIleGlyHis 83
|||||
583 TGCCAGAGAAATGGATCTCTCTCTTCTTGCTTGCTGTTCAATAGGCCATA 632
83 ..... 83
633 TAACTCAATTCTATGGGATTAATAGGCCAATATACCAATCTCTGAGGCTT 682
83 ..... 83
683 CTGCATTTCTATGTTATGCCGGTGTTAATGTGACGGTATGACTACTTC 732
84 ..... AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnA 97
733 ATGGAAGAAAGAAATCGAATGTGGAGAAAGAACCGTCTTCTTCTATGGAACA 782
97 sNHisCysG1leGlyThrAspLeuAsnArgAsnPheAlaSerIysHisTrp 113
|||||
783 ATCATTTGCGACGAAACAGACTGGAATAGGAAGACTTGTCTTCCAAACACTGG 832
114 CysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLe 130
|||||
833 TCGTAGGAAGAGTGCATCCAGTTCCATCATCTCGGAACCTACTGTGCACT 882
130 uTyrProGluSerGluProGluValAlaValAlaSerPheLeuArgA 147
|||||
883 TTATCTCTGAGTCAGAACCGAAGTGAAGCACTGGCTAGTTCTTGAGAA 932
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerIysSerGln 163
|||||
933 GAAATATTCACCAAGATTAAAGCATATCATGATCATTCATCTCCACG 982
164 HisIleValPheProTyrSerTyrThrArgSerIysSerIysAspHisG1 180
|||||
983 CATATAGAGTGTTCATATTCCTATACAGAACTAAAGAAAGACCATGA 1032
180 uGluLeuSerLeuValAlaSerGluAlaValAlaArgAla1leGluIysTrp 197
|||||
1033 GGAAGTCTCTAGTAGCCAGTGAAGCACTGCTCTATTGAGAAACTA 1082
197 eRlysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
|||||
1083 GTAAATAATACCAAGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 1132
214 AlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleIysTyrSe 230
|||||
1133 GCTCCTGGAGTGGGAGCATTTGATGATTTGGGCATCAATAATATTC 1182
230 r.Phe..... 231
1183 GTTTCACAAATTGAACCTCGAGATACGGGCACATACGATTTCTGTCGCCG 1232
232 .....ThrSerAsnProProValGluIysLeuLeuProLeuSerLeuI 246
1233 AGCGTTACATCAAAACCCACTGTAGAGAAAGCTTTTGGCGGTGTCTTAA 1282
246 s 246
1283 A 1283
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-430-787A-2
seq_documentation_block:
; Sequence 2, Application US/08430787A
; Patent No. 5593674

```

```

; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Raton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Gentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 Inch, 360 KB floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Gentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-430-787A-2

alignment_scores:
    Quality: 1272.00      Length: 301
    Ratio: 5.171          Gaps: 3
    Percent Similarity: 81.728    Percent Identity: 81.395

alignment_block:
US-09-980-881-4 x US-08-430-787A-2 ..

Align seg 1/1 to: US-08-430-787A-2 from: 1 to: 1749

1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGlnIleTyrSe 17
|||||
383 GCCTCCCGCATGCTACTATGAACAGATATCACTACATAATGAATCTATTC 432
17 tTrrPleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
|||||
433 TTGGATAGAAATTATTAAGTGAAGCATCCGATATGCTTACAAAATTC 482
34 1sileglserserphgclulysrtyrproleutyralleulysvalser 50
|||||
483 ACATTTGGATCTCATTTTGAGAAAGTACCCACTATATGTTTAAAGGTTTCT 532
51 G1LYSG1UG1nThrAlaLysAsnAla1leTrrPleAspCysGlyIleH1 67
|||||
533 GGAAGAAAGACAAACAGCCAAAATGCCATATGGATTGACGTGGAAATCCA 582

```

```

1  NUMBER OF SEQUENCES:      8
2  CORRESPONDENCE ADDRESS:
3  ADDRESSEE: Cooper & Dunham LLP
4  STREET: 1185 Avenue of the Americas
5  CITY: New York
6  STATE: New York
7  COUNTRY: U.S.A.
8  ZIP: 10036
9
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: PatentIn Release #1.0, Version #1.30
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/782,760
17 FILING DATE: 13-JAN-1997
18 CLASSIFICATION: 435
19
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: 08/378,233
22 FILING DATE: 25-JAN-1995
23 ATTORNEY/AGENT INFORMATION:
24 NAME: White, John P.
25 REGISTRATION NUMBER: 28,678
26 REFERENCE/DOCKET NUMBER: 0336/43847
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE: (212) 278-0400
29 TELEFAX: (212) 391-0525
30 INFORMATION FOR SEQ ID NO: 5:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 927 base pairs
33 TYPE: nucleic acid
34 STRANDEDNESS: single
35 TOPOLOGY: linear
36 MOLECULE TYPE: cDNA
37 HYPOTHEtical: NO
38 ANTI-SENSE: NO
39
40 NAME/KEY: CDS
41 LOCATION: 1..927
42
43 US-08-782-760-5
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
```

```

248 CTGTCCGTAACCTAATATCAAGAGATCCACATGAAACAGCTTCTAGATGAA 297
82 ..... GlyHis ..... 83
298 CTGAGATTCTATGTTCTGCTGCTGCTGCAACATTTGATGGCTATCTACAC 347
84 ..... AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsn 97
348 CTGAGACTAAGACAGAAATGTGGAGAAAACCCGCTTACTATGCTGGAA 397
97 snHisCysIleGlyThrAspLeuAsnArgAsnArgPheAlaSerLysHisTrp 113
398 GTTCCCTGCTGGGTGAGACCCCAACAGGAATTT...AATGCTGGCTGG 444
114 CysGluGluGlyAlaSerSerSerSerSerSerGluThrTyrCysGlyLe 130
445 TGTGAAGTGGAGACTTCTGGAGTCCCTCTCTGAACCTTACTGTGAGCC 494
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
495 AGCCCCAGAGTGTGAAAAGAGCAAAAGCCCTGGCAGATTTCATCCGCA 544
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
545 ACAACCTCTCCACCATCAAGGCTTACCTGACCATCATCATCATCATCATC 594
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGln 180
595 ATGATGCTCTACCCCTTACTCTATGACATCAAAAGTCCCTGAGAGATGTA 644
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysTrp 197
645 GGAATGTAATGCCCTGGTGTAAGGTGGCGCAAGAGACTT...GCCACATC 691
197 eLysAsnThrArgTyrThrArgHisGlyHisGlySerGluThrLeuTyrLeu 213
692 TGCATGGGACCAAGTACATATGCCCCAGAGCTACCAACAATCTATCTCT 741
214 AlaProGluGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTrp 230
742 GCTGCTGGGGGATCTGAGACGACTGCTTATGATCAGGGAATCAATATATTC 791
230 PheThr 232
792 CTTTACC 798

seq_name: /cgn2.6/prodata/1/lna/PCTUS_COWB.seq:PCR-US96-00995-5
seq_documentation_block:
; Sequence 5, Application PC/TUS9600995
; GENERAL INFORMATION:
; APPLICANT: Bio-Technology General Corp.
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00995
; FILING DATE: 25-JAN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0336/43847-A-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 927 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..927
; PCT-US96-00995-5

alignment_scores:
Quality: 610.00 Length: 269
Ratio: 3.389 Gaps: 5
Percent similarity: 66.914 Percent identity: 44.981

alignment_block:
US-09-980-881-4 x PCT-US96-00995-5 ..
Align seg 1/1 to: PCT-US96-00995-5 from: 1 to: 927

1 AlaSerAlaSerTyrTyrGluGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
1 GCACATGACACAGCTACACCAAGTACCAACAACTGGAAAGATGAGAGC 50
17 rTriPleGluPheIleThrGluAlaGHisProAspMetLeuThrLysIleH 34
51 GTGATTCACACAGTGGCCACTGTGATATCCAGACCTTCTCATCAGAGCG 100
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuValSer 50
101 TCATTGGAAACCATTTTGAAGAGCTACATGATGCTCTCAAGATTT... 147
51 GlyLysGluGlnThrAlaLysAsnAlaIleTrrPleAspCysGlyIleH 67
148 GGTAAACCTAGACCGAATAGCTCCCATCTTCATGATGCTGGTTGCCA 197
67 sAlaArgLutrrPleSerProAlaPheCysLeuTrpPhele..... 81
198 TGCAGAGAGTGGATTCTCTGCAATCTGTACGTGGTTGTGAGAGAG 247
81 ..... GlyHis ..... 83
248 CTGTCCGTAACCTAATATCAAGATCCACATGAAACAGCTTCTAGATGAA 297
82 ..... GlyHis ..... 83
298 CTGATTTCTATGTTCTGCTGCTGTCACATGATGATGCTGCTGACAC 347
84 ..... AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsn 97
348 CTGACACTAAGACAGAAATGTGGAGAAAACCCGCTTACTATGCTGGAA 397
97 snHisCysIleGlyThrAspLeuAsnArgAsnArgPheAlaSerLysHisTrp 113
398 GTTCCCTGCTGGGTGAGACCCCAACAGGAATTT...AATGCTGGCTGG 444
114 CysGluGluGlyAlaSerSerSerSerSerSerGluThrTyrCysGlyLe 130
445 TGTGAAGTGGAGCTTCTGGAGTCCCTCTCTGAACCTTACTGTGAGCC 494
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147

```



```

495 AGCCCGAGCTCTGAAAGAGACAAAGGCCCTGGCAGATTTCATCCGCA 544
147 rgsnllleasnlllelysalatyrilsermethisertysergin 163
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545 ACAACCTCTCCACATCAAGGCTACCTACCATCATCATCATCATCAAG 594
164 HsIlleValPheprotyrSerlyrThrArgserlySerlyspHisGI 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
595 ANGATGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 644
180 uclueuserleuValAlaserGIuValArgAlaIleGIuLysThrS 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
645 GGAATTGATGCCCTGTGTGAAAGGTGCGCAAGAGACTT...GCCACTC 691
197 eLysAsnThrArgTyrThrHisGIyHisGIySerGIuThrLeuTyrLeu 213
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
692 TGCATGGCACAAGTACATATATGATGATGATGATGATGATGATGAT 741
214 AlaProGIyGIyGIyAspAspTrrPleTyrAspLeuGIyIleLysTyrSe 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 GCTGTGGGGGATGTGACGACTGTCTTATGATCAGGGAATCAAAATTC 791
230 rPheThr 232
|||||
792 CTTTACC 798

```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:us-08-696-139-3

seq_documentation_block:

```

: Sequence 3, Application US/08696139
: Patent No. 5672496
: GENERAL INFORMATION:
: APPLICANT: Fayerman, Jeffrey T.
: APPLICANT: Greenen, David P.
: APPLICANT: Hersberger, Charles L.
: APPLICANT: Larson, Jeffrey L.
: APPLICANT: Sterner, Jane L.
: APPLICANT: Zhang, Haichao
: TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
: TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESS: Eli Lilly and Company
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: United States of America
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/696,139
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/153,258
: FILING DATE: 16-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Gaylo, Paul J.
: REGISTRATION NUMBER: 36,808
: REFERENCE/DOCKET NUMBER: X-8681
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317) 276-0756
: TELEFAX: (317) 276-3861
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 921 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```

```

; MOLECULE TYPE: CDNA
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..921
US-08-696-139-3

```

```

alignment_scores:
  Quality: 582.00      Length: 264
  Ratio: 3.326
  Percent Similarity: 66.288      Percent Identity: 44.318
alignment_block:
US-09-980-881-4 x US-08-696-139-3
Align seg 1/1 to: US-08-696-139-3 from: 1 to: 921

```

```

6  TyrGIuInTyrHisSerLeuAsnGIuIleTyrSerTrpIleGIuPheI 22
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16  TATGAGAAGTACAAACAACTGGAAACGATCGAGGCTTGACTTAAGCAAGT 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22  eThrGIuArgHisProAspMetLeuThrLysIleHisIleGIySerSerP 39
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66  CACCAGTGAATTCACAGACCTCATCTCGCACAGCCATCGGAATCATAT 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39  heGIuLysTyrProLeuTyrValLeuLysValSerGIyLysGIuInThr 55
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116  TTTTAGAACAATATATATACCTCCTCAAGGT...GGCAAACTGGAGCA 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56  AlAlAsnAlaIleTrpIleAspCysGIyIleHisAlaArgGIuTrpI 72
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163  AATTAAGCTCGCCATTTTCATGAGACTGTGGTTCCATCGCAGAAATGAT 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72  eSerProAlaPheCysLeuTrpPheIle..... 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213  TTCCATGCAATTTTGGCCAGTGTGTGTGACAGAGCGTGTCTCACTATG 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81  ..... 81
263  GATATGAGACTCATGACAGAAATTCCTCAACAAGCTATTATATGTC 312
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
82  .....GlyHis.....AsnAr 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313  TTGCGTGTGCTCAATATATGATGGCTCATCTACACCTGACCAAGACCG 362
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85  gMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGIyT 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363  AATGTGAGAAAGACCCGCTCTACCAATGCTGGAACCTGCACTTGGCA 412
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102  hrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGIuGIuAla 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
413  CAGACCCCAACAGAAATTTT...GATGCTGGGTGGTCACAACTGAGGCC 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119  SerSerSerSerCysSerGIuThrTyrCysGIyLeuTyrProGIuSerGI 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460  TCTACAGACCCCTCGATGAGACTTACTGTGAGACTCTCGCAGAGTGTGCA 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135  uProGIuValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGIu 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510  AAAAGAGACCAAGGCCCTGCTGCTGATTTTATACGCAACACTCTCTCCA 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152  leLysAlaTyrIleSerMetHisSerTyrSerGIuHisIleValPhePro 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560  TCAAGCATACCTGACATCAGTACATACACAGATGATCATCTTACCT 609
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169  TyrSerTyrThrArgSerLysSerLysAspHisGIuLeuSerLeuVal 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
610  TATTCCTATGATTAACAACCTCCCGAGAGACATCTGATGAATTAACCT 659
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185  lAlaSerGIuAlaValArgAlaIleGIuLysThrSerLysAsnThrArgT 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
660  GCCTAAGCTGCGCTGGAAGAACTT...GCTACACTGTATGACCACTAGT 706
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

310 TATGGAAGTCAACAACCTGGGAAACGATCGAGCCCTTGACTTAAGCAAGT 359
22 ethhgluwrghlspromsmellenthrysllehlsllelglsersepr 39
360 CACAGGTAAATCCAGACCTCATCTCTCCGACAGGACATCGGAACATCAT 409
39 heglulstypProleutyValleulysvalSerelylsuglunlthr 55
410 TTTTAGGAACAAATATAATACCTCCCAAGGT...GGCAACCTGGACCA 456
56 Alalysanlailetrpilleapspcsglyllehlslalarglunlrpili 72
457 AATAGCGCTGCCATTTTCATGCATCTGTGTTTCCATGCCAGAAATGAT 506
72 eSerProAlaPheCysLeuTrpPheIle..... 81
507 TTCCCATGATGATTTTCCGATGGTGTGTGAGAGAGGCTGTTCCACCTAG 556
81 ..... 81
557 GATATGAGATCAATGACAGAAATTCCTCCMACACCTAGACTTTATGTC 606
82 .....GlyHis.....AsnAr 85
607 TTGCCTGTGCTCAATATGATGGCTACATCAACCTGGACCAAGAACGG 656
85 gmettrparylgsantrgserrpneThyAlaAsnAsnHscyslllellyT 102
657 AATGTGAGAAAGACCCGCTCTACCAATGCGGAACCTACCTCATTTGCA 706
102 hrAspleuAsnArAsnPheAlaSerleThySHstrpCysglunlglulAla 118
707 CAGACCCCAACAGAAATTT...GATGCTGGGTGTGCACAACTGAGACC 753
119 SerSererSerCysSerelulThrTyrcysglulentyrProgluSerel 135
754 TCACAGACCCCTCGATGACACTTACGTGATCTGCTGCAGAGACTGCA 803
135 uprogluValylsAlaValAlaSerPheLeuArgarlnleAanglnI 152
804 AAAAGAGACCAAGCCCTGGCTGATTTTATACGCAACAACCTCTCTCCA 853
152 lelulalatyrlleSerMethlSerTySergrlnHslleValPhepro 168
854 TCAAAGCATACCTGACCATCCATCACTACAGATGATACTTACCT 903
169 TyrSerTyrrhrArgserlyserlyAspHslglunleuSerleuVal 185
904 TATTTCTATGATTACAAACATCCCCGAGAACATCCTGAGTGAATACCT 953
185 lAlaSergluAlaValArgAlaIleglulstypThSerlysanthrarGT 202
954 GGCTAAGGCTGCGGTGAAGAAGTT...GCTACACTGATAGGCCAACAGT 1000
202 yrrhrHslglulsglyserelulThreutyrlleuAlaProgllyelgyl 218
1001 ACACATACGCGCCAGGAGCTCAACAATCATCTGCTGCGGGGGGCTCT 1056
219 AspAspTrpIleTyraAspleuglyllellystysSerPheThr 232
1051 GATGACTGGGCTTATGACCAAGGATCAATATTCCTCTTACC 1092

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-860-882A-67
seq_documentation_block:
; Sequence 67, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGEITTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT

```

APPLICANT: BOYLE, FRANCIS THOMAS
APPLICANT: HENNAM, JOHN FREDERICK
APPLICANT: BLAKELY, DAVID CHARLES
APPLICANT: MARSHAM, PETER ROBERT
APPLICANT: HEATON, DAVID WILLIAM
APPLICANT: DAVIES, DAVID HOW
TITLE OF INVENTION: CHEMICAL COMPOUNDS
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESS: PILLSBURY, MADISON & SUTRO
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860, 882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 999 bases
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
US-08-860-882A-67

alignment_scores:
Quality: 567.00 Length: 269
Ratio: 3.133 Gaps: 5
Percent Similarity: 67.286 Percent Identity: 41.264

alignment_block:
US-09-980-881-4 x US-08-860-882A-67 ..

Align seg 1/1 to: US-08-860-882A-67 from: 1 to: 999

```
1 Alaserlaserlyrrgluglunlyrhisserleuasnclulelyrse 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 GCACACTGGCTCTTACGAGAGTACACACAGTGGAAACGATAGAGGC 116
17 rtrpliegulphelrthrglunrghlspasmetleuthrlyslth 34
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 TTGACTCAACAGTGGCCACATGACATCCAGCCCTCATCTCGCAGTG 166
34 lstrlelgserserphneglulystyrproleuthryvallleulysvalser 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TTYTCGAGACACATTGAGGAGCGGCTATTTACTCTCGTGAAGGTT... 213
51 Glylysgluginthralalysasnalalettrpilesapcysglylehl 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 GCGAAAGCTGACAAATAAGCCTGCATTTTCATGACGCTGTGTTCCA 263
67 salaarglutrpileserproalaphecysleutrpheile..... 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TGCCAGAGAGTGTTCCTCTCATTTCTCCAGTGTGTTAAGAGAGC 313
81 ..... 81
```

```
314 CTGTTCGTACCTATGACGTGAGATCCAGTACAGAGCTTCTCGACAG 363
82 .....glyhis..... 83
364 TTAGACTTTATGTCCTGCTGCTGCTCAATATTTGATGAGCTACATCTAC 413
84 .....AsnArgMetTrpArgLysAsnArgSerPheTrpAlaAsnA 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 CTGACCAACAGCCGATTTGAGAAAGACTCGTCACCACTCATCTGAT 463
97 snhiscysllieglythrspgleuasnargasnphlealaserlyshstip 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 CTAGCTGATTTGGCAGACACCCCAAGAAATTT...GATGCTGTTGG 510
114 CysglugluclyalaserSerSerCysSerCylunthryrCysglyle 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
511 TGTGAATTTGAGAGCTCTCGAAGCCCTGTGATGAACTACTGTGAGACC 560
130 uTyrProglusergluprogluvalylsvalalaserPheleuargA 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 TGCCGACAGCTGTGAAAGAGCAAGCCCTGCTGATTTTCATCCGCA 610
147 rgasnleasnghllyalalyrlylesermethtlsserlyserglu 163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
611 ACAACCTCTCTCCATGAGCATATCTGACAACTCATCTGATCCCA 660
164 HislevalPheProTyrSerTyrTrpArgSerlyserlysasphlsl 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
661 ATGATGATCTACCTTTCATATGCTTACCAACTCGGTGAGAACATGTC 710
180 uglulserleuvalalaserglualalargalileglulysThrs 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
711 TGAATTGAATGCCCTGCTAAGCTACTGTGAAGAACTT...GCCATCAC 757
197 erlyasntrharglythrhisgllyhsglysergluthrleuyleu 213
758 TGCAGGACCAAGTACACATATGCCCCGGAGGTACACACATATTCCT 807
214 Alaprogllygllyaspasrtrpilletraspplengllyllyltyrse 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
808 GCTGCTGGGGCTCTGACAGCTGGCTTATGACCAAGCATCATATTC 857
230 rPheThr 232
|||||
858 CTTCAC 864
```

seq_name: /gsn2_6/ptodata/1/lna/5b_COMB.seq:US-08-860-882A-64

seq_documentation_block:

```
; Sequence 64, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGSELTRE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESS: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
```

```

; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1053 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-860-882A-64

```

```

alignment_scores:
    Quality: 567.00      Length: 269
    Ratio: 3.133        Gaps: 5
    Percent Similarity: 67.286    Percent Identity: 41.264

```

```

alignment_block:
US-09-980-881-4 x US-08-860-882A-64 ..

```

```

Align seg 1/1 to: US-08-860-882A-64 from: 1 to: 1053

```

```

1  AlaseralaserTYTYRGLUGLINTYRHISSErLeuAnGluileTyse 17
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67  GCACTGCTCACTCTTACGACAGATGACAAAGTGGAAAGCATGAGACC 116
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
17  TTPrlleGluPheilethrgluArgHISProAspMetleuthrlyleH 34
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
117  TTGGACTCAACAGTCCGCTGAGATCCAGCCCTCATCTCTGCGCACTG 166
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
34  IsileGlySerSerPheGluLysYrProLeuYrValleuLysAlser 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
167  TTATCGGAACACATTTGAGGGAGCGCTATTACTCTCTGAAGGT... 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
51  GLYysgluglInthrAlaLysAsnAlaileTrrPrlleAspCysGlyleH1 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
214  GGCAAGCTGACAAATAAGCTGCCATTTTCATGAGACTGCTGTTCCA 263
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
67  sAlarGluTrpPrlleSerProAlaPheCysLeuTrPheile..... 81
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
264  TGCCAGAGAGTGTCTCTCTGATCTGCGAGTGTGTGTAAGAGAG 313
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
81  ..... 81
314  CTGTTCGACTATGAGAGTGAATCCAGTGAAGAGCTTCTGACAGAG 363
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
82  .....glyHis..... 83
364  TTAGACTTTTATGTCTGCTGCTGCTCAATATGATGGCTACATCACTA 413
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
84  .....AsnArgMetTrpArgLysAsnArgSerPheYrAlaAna 97
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
414  CTGACAGCAAGAGCGGATTTGGAGAAAGACGCTGCCACCATCTGAT 463
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
97  snHISysileGlyThrAspLeuAsnArgPheAlaSerLysHisTyr 113
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
464  CTAGCTGATTTGACAGACAGACCCCAAGAAATTT...GATGCTGGTGG 510
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
114  CysgluglInglYalaserSerSerCysSerGluThrTyrcysGlyle 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
511  TGTGAATTTGAGGCTCTCGAAACCCCTGTGATGAACCTTACTGTGAGC 560

```

```

130  uTYrProGluSerGluProGluValLysAlaValAlaserPheLeuArg 147
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
561  TGGCCGAGAGTGTGAAGAGAGACCAAGGCCCTGGCTATTTCATCCGCA 610
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
147  rGAsnIleAsnGlnIleLysAlaYrTlleSerMetHisSerYrSerGln 163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
611  ACAAACTCTTCCATCAAGGCATATGACAAATCCACTGACATGCCAA 660
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
164  HsIleValPheProTYrSerTYrThrAspSerLysSerLysAspHisG1 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
661  ATGATGATCTACCTTACTATGCTTACAAACTGCGAGAACAAATGC 710
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
180  uGluLeuSerLeuValAlaserGluAlaValArgAlaileGluLysTrp 197
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
711  TGAGTTGAATGCCCTGGCTGAAGACATGTAAGAACTT...GCCCTAC 757
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
197  eLYsAsnThrArgTYrThrHisGlyHisGlySerGluThrleuTYrleu 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
758  TGCACGGGACCAAGTACATATGCCCCGGAGCTACAAACATCTATCT 807
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
214  AlAProGlyGLYLYAspAspTrpIleTYrAspLeuGlyIleLysTyse 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
808  GCTGCTGGGGGCTCTGACGACTGGCTTATGACCAAGAAATCAGATATTC 857
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
230  rPheThr 232
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
858  CTCAC 864

```

```

seq_name: /cgn2_6/prodata/1/ina/5b_comb.seq:US-08-860-882A-56

```

```

seq_documentation_block:
; Sequence 56, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSER: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:

```



```

1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 GCACAGGACACAGTATGAGAAAGTACAAACAGTGGAAACATAGAGGC 401
17 TTrpIleGluPheIleThrGluArgHisProAspMetLeuThrIleH 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
402 TTGGAGCTCAACAAGTCGCCACTGAGAAATCCACCCCTCATCTCCGAGTG 451
34 IsIleGlySerSerPheGluTyrTyrProLeuTyrValLeuIysValSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
452 TTATCGAAGCACATTTGAGGAGCGGGTATTACTCTCGAAGTT... 498
51 GlyIysGluGlnThrAlaIysAsnAlaIleTrrPleAspCysGlyIleH 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
499 GGCAGAGCTGACAAATTAAGCCCTCATTTTCATGAGACTGGTTGCCA 548
67 sAlaArgGluTrrPleSerProAlaPheCysLeuTrrPheIle..... 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
549 TCCACAGACAGTGGATTCTCTCCATTCCTGCCAGTGGTTTGTAGAGAGG 598
81 ..... 81
599 CTGTTCTGACTATGAGAGTGATCAAGTACAGAGCTTCTGACAAG 648
82 .....GlyHis..... 83
649 TTAGACTTTATGCTCCTGCTGCTCATATTTGATGCTACATCTACAC 698
84 .....AsnArgMetTrrPArgIysAsnArgSerPheTrrAlaAsnA 97
699 CTGACGACCAAGCCGATTTTGAGAAACACTCGCTCCACCCCTACTGGAT 748
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerIleHisTrr 113
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
749 CTAGCTGATTTGGCACAGACCCCAACAGAAATTT...GATCGTGGTGG 795
114 CysGluGluGlyAlaSerSerSerSerSerSerGluTrrTyrCysGlyLe 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
796 TGTGAATTGGAGCCTCTCGAAACCCCTGTGATGAACTTACTGTGGACC 845
130 uTyrProGluSerGluProGluValAlaValAlaSerPheLeuArgA 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
846 TGCCGCAAGCTGTGAAGAAGACACAGCCCTGGCTGATTTTCATCCCA 895
147 rGAsnIleAsnGlnIleIysAlaTrrIleSerMetHisSerTyrSerGln 163
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
896 ACAAACTCTCTCATCAAGCATATCTGACAATCCACTCGTACTCCCAA 945
164 HisIleValPheProTyrSerTyrTrrArgSerIysSerIysAspHisG 180
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 ATGATGATCTACCTTACTCATATGCTTACAAACTCGGTGAGAACAAATGC 995
180 uGluLeuSerLeuValAlaSerGluAlaValAlaIleGluIysTrrs 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
996 TGAGTTGAATCCCTGCTAAAGCTACGTGAAGAACTT...GCCTTAC 1042
197 eRlyAsnThrArgTrrTrrHisIleGlySerGlyIleThrLeuTrrLeu 213
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1043 TGCACGGACCAAGTACACATATGCGCGGAGCTACAAACAATCATCT 1092
214 AlaProGlyGlyIysAspAspTrrIleTrrAspLeuGlyIleIysTrrSe 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1093 GCTCTGGGGGCTCTGACGACTGGCTTATGACCAAGAAATCAATATTC 1142
230 rPheThr 232
|||||
1143 CTTCACC 1149
seq_name: /cgn2_6/ptodata/1/lna/5B_COMB_seq: US-08-860-882A-74
seq_documentation_block:
; sequence 74, Application US/08860882A
; Patent No. 5985281

```

```

; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOU, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1059 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-860-882A-74

alignment_scores:
Quality: 560.00 Length: 269
Ratio: 3.111 Gaps: 5
Percent Similarity: 66.914 Percent Identity: 40.892

alignment_block:
US-09-980-881-4 x US-08-860-882A-74 ..
Align seg 1/1 to: US-08-860-882A-74 from: 1 to: 1059

1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 GCACTGCTACTCTTACGAGAGTACACAAAGTGGAAAGATAGAGGC 116
17 TTrpIleGluPheIleThrGluArgHisProAspMetLeuThrIleH 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 TTGACACTCAACAAGTCGCCACTGAGAAATCCACCCCTCATCTCCGAGTG 166
34 IsIleGlySerSerPheGluTyrTyrProLeuTyrValLeuIysValSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 TTATCGAAGCACATTTGAGGAGCGGGTATTACTCTCGAAGTT... 213
51 GlyIysGluGlnThrAlaIysAsnAlaIleTrrPleAspCysGlyIleH 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
214 GGCAGAGCTGACAAATTAAGCCCTCATTTTCATGAGACTGGTTGCCA 263
67 sAlaArgGluTrrPleSerProAlaPheCysLeuTrrPheIle..... 81

```

```

|||||
264 TGCAGAGAGATGATTTCTCTCCATTTCCAGTGTGTTTAAGAGAG 313
81 ..... 81
314 CTGTTGTAACCTATGAGAGTGAATCCAGTACAGAGCTTTCGACAA 363
82 ..... 83
364 TTAGACTTTATGTCCTGCTGCTCAATATTTGATGATGACATCTAC 413
84 .....
414 CTGGACCAAGAGCCGATTTTGAGAAAGACTCGTCCACCCATCTGAT 463
97 snhiscysilleglythrAspleuasnargasnphelaaserlysh1strp 113
:: |||||
464 CTAGCTGCAATGGCAGACAGACCCACAGAAATTTT...CATGCTGTTGG 510
114 CysglugluglyAlaserSerSerSerSerSerSerSerSerSerSer 130
|||||
511 TGTGAATTTGAGCCCTCGAACCCTGATGATGAATTAATCTGAGACC 560
130 wtyrProgluSerGluProgluValAlaValAlaValAlaSerPheLeuArg 147
|||||
561 TGCCGAGAGTGTGAAGAGAGCAAGCCCTGCTGATTTATCCGCA 610
147 rgsanlleasnnglnlleysAlaTyrIleSerMetHisSerTyrSerGln 163
:: |||||
611 ACAACTCTCTTCATCAAGGCAATATGATCATCCATCTGATCTCCCA 660
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 180
:: |||||
661 ATGATGATCTACCTTACATATGCTTACAAACTCGGTGAGAAACATGC 710
180 uGluLeuSerLeuValAlaSerGluValAlaValAlaIleGluLysThr 197
|||||
711 TGAGTTGAATGCCCTGCTAAAGCTACTGTGAAGAACTT...GCCTCAC 757
197 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
:: |||||
758 TGCAGCGACCAAGTACACATATGCGCCGAGACTTACAAACATCTATCC 807
214 AlaProGlyGlyLysAspAspTrrPleTyrAspLeuGlyLysTyrSer 230
|||
808 GCGTGGGGGGCTTAAAGACTGGCTTATGACCAAGAAATCAGATATTC 857
230 rPheThr 232
|||||
858 CTTCACC 864

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-860-882A-77

seq_documentation_block:

Sequence 77, Application US/08860882A

Patent No. 5985281

GENERAL INFORMATION:

APPLICANT: TAYLORSON, CHRISTOPHER JOHN

APPLICANT: EGGELE, HENDRIKUS JOHANNES

APPLICANT: TARRAGONA-FIOL, ANTONIO

APPLICANT: RABIN, BRIAN ROBERT

APPLICANT: BOYLE, FRANCIS THOMAS

APPLICANT: HENNEM, JOHN FREDERICK

APPLICANT: BLAKELY, DAVID CHARLES

APPLICANT: MARSHAM, PETER ROBERT

APPLICANT: HEATON, DAVID WILLIAM

TITLE OF INVENTION: CHEMICAL COMPOUNDS

NUMBER OF SEQUENCES: 77

CORRESPONDENCE ADDRESS:

ADDRESSEE: PILLSBURY, MADISON & SUTRO

STREET: 1100 NEW YORK AVENUE, N.W.

CITY: WASHINGTON

```

STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 1059 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-860-882A-77

alignment_scores:
Quality: 559.00 Length: 269
Ratio: 3.106 Gaps: 5
Percent Similarity: 66.914 Percent Identity: 40.892

alignment_block:
US-09-980-881-4 x US-08-860-882A-77 ..
Align seq 1/1 to: US-08-860-882A-77 from: 1 to: 1059

1 AlaserAlaserTyrTyrGluGluThrHisSerLeuAsnGluIleTyrSer 17
|||||
67 GCACCTGTCTACTTTCAGCAATACACAGAGTGGGAACGATAGAGGC 116
117 TTGACCTACCAAGTCCCACTGAGATCCAGCCCTCATCTCTGCAAGTG 166
17 rTrrPleGluPheIleThrGluArgHisProAspMetLeuThrLysIle 34
|||
117 TTGACCTACCAAGTCCCACTGAGATCCAGCCCTCATCTCTGCAAGTG 166
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
|||||
167 TTATCGGAGACCAATTTGAGGAGCGCGCTATTTACTCTGAAAGTT... 213
51 GlyLysGluGluThrAlaLysAsnAlaIleTrrPleAspCysGlyIleH 67
|||||
214 GGCNAAGCTGACAAATTAAGCTGCGCATTTTCATGAGCTGTGGTTCCA 263
67 sAlaArgGluTrrPleSerProAlaPheCysLeuTrrPheIle..... 81
|||||
264 TGCAGAGAGTGAATTTCTCTGATCTGCGCAATGAGTTGTTAAGAGAG 313
81 ..... 81
314 CTGTTGTAACCTATGAGAGTGAATCCAGTACAGAGCTTTCGACAAAG 363
82 ..... 83
364 TTAGACTTTATGTCCTGCTGCTCAATATTTGATGATGACATCTAC 413
84 .....
414 CTGGACCAAGAGCCGATTTTGAGAAAGACTCGTCCACCCATCTGAT 463
97 snhiscysilleglythrAspleuasnargasnphelaaserlysh1strp 113
:: |||||

```


OM of: US-09-980-881-4 to: EST:* out_format : pfs
Date: Sep 18, 2002 5:41 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+ .p2n.model -DEV-x1h
-O=/cgn2_1/USPRO.spool/US09980881/runat_16092002_140128_10623/app-query.fasta_1.1126
-DB-EST -OPM-fasta+ -SUFFIX=first -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OCGAPOP=4.500
-OCGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCGAPOP=6.000
-FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELDP=6.000
-DELEXT=7.000 -START=1 -MATRIX-biosum62 -TRANS-numan40.cdi
-LIST=45 -DOCMLOCAL -200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTEXT=pfs -NORM-ext -HEADSIZE=500
-MILEN=0 -MAXLEN=200000000 -USER=US09980881@cgn1.1.3434
-NCPU=6 -ICPU=3 -LONGLOC -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLXRY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-4

Query length: 246

Database: EST:*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 3115.820000

score_list:	Strd Orig	ZScore	EScore	len	Documentation
gb_hnc:AK004045	+ 1112.00	2190.10	9.6e-113	1446	AK004045 Mus musculus 18 days
gb_est2:BG569281	+ 978.00	1928.94	3.4e-98	852	BG569281 602588573P1 NIH_MGC_76
gb_est1:AV651709	+ 831.50	1639.14	4.7e-82	688	AV651709 AV651709 GIC Homo sapi
gb_est2:BG567264	+ 828.50	1631.29	1.3e-81	806	BG567264 602589745P1 NIH_MGC_76
gb_est1:AL536068	+ 812.50	1603.33	4.6e-80	575	AL536068 AL536068 LTI_FLO13_FBI
gb_est1:AW413090	+ 784.50	1542.34	1.2e-76	889	AL536069 AL536069 LTI_FLO13_FBI
gb_est1:AI259599	+ 688.50	1373.17	3.1e-67	723	AW413090 uq52c03.x1 Sugano mus
gb_est1:AI259599	+ 685.50	1346.46	9.4e-66	772	AI259599 u183d11.x1 Sugano mus
gb_est2:BF384332	+ 659.50	1329.96	8.0e-65	692	BF384332 60204670T1 NCI_CGAP_1
gb_est1:AI255337	+ 659.50	1294.96	7.0e-63	747	AI255337 u194a04.x1 Sugano mus
gb_est2:BI26210	+ 656.50	1289.33	1.4e-62	713	BI264210 602958724P1 NCI_CGAP_1
gb_est2:BG642020	+ 649.00	1273.74	8.2e-62	645	BG642020 p911c.pk002.n16 Chikve
gb_est2:BG641865	+ 631.00	1241.78	4.6e-60	546	BI1865 yf02b04.r1 Soares fetal
gb_est1:AI595707	+ 630.00	1235.96	1.3e-59	756	AI595707 u168d08.x1 Sugano mus
gb_est2:BG565100	+ 624.00	1224.13	6.1e-59	747	BG565100 602583720P1 NIH_MGC_76
gb_est1:AV651204	+ 607.00	1198.22	1.7e-57	378	AV651204 AV651204 GIC Homo sapi
gb_hnc:AK008678	+ 607.00	1186.30	7.9e-57	1041	AK008678 Mus musculus adult ma
gb_est2:BI144644	+ 606.00	1186.65	7.5e-57	853	BI144644 602909332P1 NCI_CGAP_1
gb_hnc:AK003061	+ 577.50	1123.72	2.4e-53	1429	AK003061 Mus musculus adult ma
gb_est2:BG216556	+ 561.50	1098.00	6.5e-52	843	BG216526 RSM36106 Athersys RAGH
gb_est1:AB030473	+ 561.00	1093.08	1.2e-51	1176	AB030473 AB030473 Mus musculus
gb_est1:AA244760	+ 560.50	1102.97	1.4e-52	466	AA244760 mx05d08.r1 Soares mus
gb_est2:BI064948	+ 547.00	1072.75	1.7e-50	616	BI064948 p911n.pk003.a6 normal1
gb_est2:BF236333	+ 542.00	1056.82	1.3e-49	1022	BI236333 60298408P1 NCI_CGAP_1
gb_est1:AB030474	+ 528.00	1034.71	2.2e-48	623	AB030474 AB030474 Mus musculus
gb_est1:AI048167	+ 523.00	1017.78	1.9e-47	1125	AI048167 u671b12.x1 Sugano mus
gb_est1:AA968389	+ 506.50	991.42	5.6e-46	644	AA968389 uc71b07.x1 Sugano mus
gb_est1:AI786705	+ 504.50	988.56	8.1e-46	585	AI786705 u156a06.x1 Sugano mus
gb_est2:BI26210	+ 500.00	984.55	1.4e-45	384	AV683230 AV683230 GKB Homo sapi
gb_est1:AV659396	+ 483.50	941.13	3.6e-43	936	AL569396 AL569396 LTI_FLO02_PLI
gb_est1:AV652344	+ 481.50	947.77	1.5e-43	379	AV652344 AV652344 GIC Homo sapi
gb_hnc:AK003088	+ 480.50	930.74	1.4e-42	1361	AK003088 Mus musculus adult ma
gb_hnc:AK015255	+ 473.50	912.44	1.4e-41	1966	AK015255 Mus musculus adult ma
gb_est2:BG616436	+ 466.50	910.00	1.9e-41	738	BG616436 60261618P1 NIH_MGC_76
gb_est1:AI196759	+ 465.50	911.13	2.0e-41	566	AI196759 u166b09.x1 Sugano mus
gb_est1:AI527277	+ 464.50	909.77	1.7e-41	536	AI527277 u156g09.x1 Sugano mus
gb_est2:Y73411	+ 461.50	910.36	1.9e-41	788	Y73411 yc034h03.r1 Stratae 13
gb_est2:BG618629	+ 461.00	898.26	8.7e-41	788	BG618629 602646064P1 NIH_MGC_76
gb_est2:BG618239	+ 456.00	889.26	2.8e-40	725	BG618239 602645967P1 NIH_MGC_76

gb_est2:BG566051	+ 445.00	866.96	4.8e-39	747	BG566051 602582552P1 NIH_MGC
gb_est2:W89198	- 438.50	856.70	1.8e-38	593	W89198 z169f104.s1 Soares_fet
gb_est2:BI766188	+ 433.00	843.14	1.0e-37	739	BI766188 603052820P1 NIH_MGC
gb_est2:BI792659	+ 431.00	841.64	1.2e-37	598	BI792659 lc39c12.y1 Melton N
gb_est2:BI712255	+ 430.00	840.00	1.5e-37	580	BI712255 1b77e07.y4 Amplifite

seq_name: gb_hnc:AK004045

seq_documentation_block:

LOCUS AK004045 1446 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus 18 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:1110032P04:carboxypeptidase B2 (plasma),
full insert sequence.

ACCESSION AK004045
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,
clone:11b:RIKEN full-length enriched mouse cDNA library
clone:1110032P04.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Carlincl,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PubMed 10349636

REFERENCE

AUTHORS Carlincl,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 2049374
PubMed 11042159

REFERENCE

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carlincl,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashito,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T. and Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PubMed 11076861

REFERENCE

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PubMed 11076861

REFERENCE

AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Araiawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
Carlincl,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanaoka,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirooka,T., Hoti,F.,
Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
Kato,H., Kawal,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
Kurahara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
Numata,K., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Schimml,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tasaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamana,T.,
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.

TITLE

Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Evolution Research Group, RIKEN Genomic Sciences Center (GSC),

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 852)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: UICM1580 row: b column: 19
 High quality sequence stop: 751.
 Location/Qualifiers
 1..852
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4722354"
 /clone_lib="NIH-MGC_76"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcgccctcgagcc); Site_2: SfiI (ggccatagagcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CAGGCGCATTTATGGCC-3' and 3' adaptor sequence:
 5'-ATCTAGAGCGCCGAGCGCCGACATG-dt(30)-BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH-MGC library."

FEATURES
 source
 alignment_scores:
 Quality: 978.00 Length: 260
 Ratio: 4.866 Gaps: 5
 Percent Similarity: 77.308 Percent Identity: 75.385
 alignment_block:
 US-09-980-881-4 x BG569281 ..
 Align seg 1/1 to: BG569281 from: 1 to: 852
 BASE COUNT 251 a 177 c 187 g 237 t
 ORIGIN

43 ProlenTtValleulysValSerGlyLysGluGlnThrAlaLysAsnAl 59
 |||||||
 2 CCACTCTATGTTTAAAGGTTTCTGGAAGAACACACACCAAAATATGC 51
 |||||||
 59 alleTtPileaspGlyLleHisAlaArgLutPileSerProAlar 76
 |||||||
 52 CAATGATGATGACTGTGGAATCCATGCCAAGAATGATCTCTCCGCTT 101
 |||||||
 76 heCysLeuTtPheileGlyHis..... 83
 |||||||
 102 TCGGCTTGCTGTCATAGGCCATATACTCAATTCTATGGGATATAGG 151
 |||||||
 83 83
 152 CAATATACCAATCTCGAGGCTTGATTTCTATGTATGCCAGTGT 201
 |||||||
 84AsnArgMetTtP Arg 88
 |||||||
 202 TAATGTGATGTTTATGACTACTCATGGAAGAAAGATGATGATGAGCA 251
 |||||||
 89 LysAsnAlrSerPheTtAlaAsnAlHisCysAlleGlyThrAspLeuAs 105
 |||||||
 252 AAGAAGCGTCTTCTATGCGAACAAATCATTCATGCGAAGACAGCCGAA 301
 |||||||

105 nArgAsnPhaAlaSerLysHisTtPcysGluGluGlyAlaSerSerSers 122
 |||||||
 302 TAGGAACCTTCTCTCCAAACACTGGTGTGAGAGAGTGCATCCAGTCTCT 351
 |||||||
 122 eCysSerGluTtThrTtCysGlyLeuTtPProGluSerGluProGluVal 138
 |||||||
 352 CATGCTCGAAGAACTACTGCTGACTTATCTCGATGAGCAACGAAATGTG 401
 |||||||
 139 LysAlaValAlaSerPheLeuArgArgAsnAlleAsnGlnLleLysAlaTy 155
 |||||||
 402 AAGCAGATGGCTACTTCTTCTAGAGAAATATCAACCGATTAAGCATTA 451
 |||||||
 155 rIleSerMetHisSerTtYSerGlnHisIleValPheProTtYSerTtY 172
 |||||||
 452 CATCAGCATGATCATCATCTCCAGCATATATGTCTTCCATATCCCTATA 501
 |||||||
 172 hArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
 |||||||
 502 CACGACAGTAAAGCAAGAACATGAGGAACGTCCTCTAGTACCGAGTCAA 551
 |||||||
 189 AlAlaValArgAlaIleGlyLysThrSerLysAsnThrArgTtYThrHisG 205
 |||||||
 552 GCAATTCGTCTATGAGAAATATGTAATAATCCAGGTTFACACATGC 601
 |||||||
 205 yHisGlySerGluTtThrLeuTtYLeuAlaProGlyGlyAlaSpasPrtP 222
 |||||||
 602 GCATGGCTCAGAAACCTATACCTAGCTCCGAGAGTGGGACGATTTGA 651
 |||||||
 222 leTtYrAspLeuGlyLleLysTtYSerPheThr..... 232
 |||||||
 652 TCTATGATTTGGGCGATCAGATATTCGTTACAAATGGAACCTTGAGATA 701
 |||||||
 233SerAsn..ProProVa 237
 |||||||
 702 CGGCGACATTAACGATTCCTGCTCCGAGAGCTTACATCAACACCACTGT 751
 |||||||
 237 lGluLysLeuLeuProLeuSerLeu 245
 |||||||
 752 AGAGAACCTTTGGCGGCTGCTCTTA 776
 |||||||

seq_name: gb_est1:AV651709
 seq_documentation_block:
 LOCUS AV651709 688 bp mRNA linear EST 15-JAN-2002
 DEFINITION AV651709 GLC Homo sapiens cDNA clone GLCSP10 3', mRNA sequence.
 ACCESSION AV651709
 VERSION AV651709.1 GI:98727223
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 688)
 AUTHORS Xu,X., Huang,D., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
 Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
 Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
 Hu,G., Gu,J., Chen,Z. and Han,Z.
 TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 MEDLINE 21625106
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES
 source
 1..688
 /organism="Homo sapiens"

```

      /db_xref="taxon:9606"
      /clone_lib="GIC"
      /tissue_type="corresponding non cancerous liver tissue"
      /dev_stage="Adult"
      /lab_host="SOLR"
      /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
      XhoI"
BASE COUNT      208 a      145 c      151 g      184 t
ORIGIN

Alignment_scores:
      Quality: 831.50      Length: 181
      Ratio: 5.133      Gaps: 2
      Percent Similarity: 89.503      Percent Identity: 88.950

Alignment_block:
US-09-980-881-4 x AV651709 ..

Align seg 1/1 to: AV651709 from: 1 to: 688

      84 AsnATGcttPArgLyAsnATGSerPheTyAlaAsnAsnHicys11 100
      |||||||
      46 AATCGAATGTGGAGAAAGACCGTCTTCTATGGAGCAATCATTCAT 95
      |||||||
      100 eGlyThrAspLeuAsnATGAsnPhaAlaSerLysHisTrpCysGluGlu 117
      |||||||
      96 CGGAACACACCTGAAATAGCACTTGTCTCCAAACACCTGCTGAGAAAG 145
      |||||||
      117 LyAlaSerSerSerCysSerGluThrTyrcysGlyLeuTyProGlu 133
      |||||||
      146 GTGCATCCAGTTCCTCATGCTCGGAAACCTAGTGAGACTTATCTGAG 195
      |||||||
      134 SerGluProGluValLyAlaValAlaSerPheLeuATGATGAsn145 150
      |||||||
      196 TCAGAACAGAGTAGAGAGAGAGTGTCTGTTCTGAGAGAAATATCA 245
      |||||||
      150 nGlnIleLyAlaTyrlIleSerMetHisSerLysSerGlnHisIleVal 167
      |||||||
      246 CCAGATTAAACATACATCAGCATGCTATTCATCTCCAGCATATAGTGT 295
      |||||||
      167 heProTySerTyThrArgSerLysSerLysAspHisGluGluLeuSer 183
      |||||||
      296 TTCCATATTCCTATACAGAAAGTAAACCAAGACCATGAGGACTGTCT 345
      |||||||
      184 LeuValAlaSerGluAlaValAlaArgAlaIleGluLysThrSerLysAsn 200
      |||||||
      346 CTAGAGCCAGTAGAAGCAGTTCGTCTATGTGAAAACTAGTAAATAATAC 395
      |||||||
      200 rArgTyThrHisGlyHisGlySerGluThrLeuTyrlleuAlaProGly 217
      |||||||
      396 CAGGTATACATGAGCCATGGCTCAGAAACCTTATACCTAGCTCTGAG 445
      |||||||
      217 LyGlyAspAspTrpIleTyAspLeuGlyIleLysTySer.Phe..... 231
      |||||||
      446 CTGGGAGAGATTGGATCTATGATTTGGGCATCAATATTCGTTTCAAT 495
      |||||||
      232 .....Thrs 233
      |||
      496 GAACCTGAGATACCGGACATACCGATTCTTTGTGGGAGAGCTTACAT 545
      |||
      233 eRAsnProProValAGLulysLeuLeuProLeuSerLeuLys 246
      |||||||
      546 CAACCCACCTGTTGAGAAAGCTTTGGCGCTGTCTTAAA 586
      |||||||

seq_name: gb_est2:BG567264

seq_documentation_block:
LOCUS      BG567264      806 bp      mRNA      linear      EST 10-Apr-2001
DEFINITION  602389745F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723690 5',
            mRNA sequence.
ACCESSION  BG567264

```

```

VERSION      BG567264.1      GI:13574917
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL
COMMENT      Contact: Robert Strausberg, Ph.D.
            Email: gcapbs-remail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM1583 row: j column: 11
            High quality sequence stop: 772.
            Location/Qualifiers
            1. 806
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_image="4723690"
            /clone_lib="NIH_MGC_76"
            /lab_host="DH10B (T1 phage-resistant)"
            /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
            SfiI (ggccgctggcc); Site_2: SfiI (ggccattatggcc); 5' and
            3' adaptors were used in cloning as follows: 5' adaptor
            sequence: 5'-CAGGCGCATATGACC-3' and 3' adaptor sequence:
            5'-ATTCTAGAGCGCCGAGCGCCGACATG-dn(30)BN-3' (where B = A,
            C, or G and N = A, C, G, or T). Average insert size 1.85
            kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
            by PCR. This library was enriched for full-length clones
            and was constructed by Clontech Laboratories (Palo Alto,
            CA). Note: this is a NIH-MGC Library."
BASE COUNT      239 a      158 c      162 g      247 t
ORIGIN

Alignment_scores:
      Quality: 828.50      Length: 179
      Ratio: 5.146      Gaps: 2
      Percent Similarity: 89.944      Percent Identity: 89.385

Alignment_block:
US-09-980-881-4 x BG567264 ..

Align seg 1/1 to: BG567264 from: 1 to: 806

      86 MetTPArgLyAsnATGSerPheTyAlaAsnAsnHicys11eGlyTh 102
      |||||||
      2 ATGTGAGAAAGAACCGTCTTCTATGCGAACAATCATTCATGCGAAC 51
      |||||||
      102 rAspLeuAsnATGAsnPhaAlaSerLysHisTrpCysGluGluLysAla 119
      |||||||
      52 AGACCTGAATAGCACTTGTCTCCAAACACTGCTGAGAGAGTGCAT 101
      |||||||
      119 eRserSerSerCysSerGluThrTyrcysGlyLeuTyProGluSerL 135
      |||||||
      102 CCAGTTCCTCATGCTCGGAAACCTAGCTGACCTTATCTGAGTCAAG 151
      |||||||
      136 ProGluValLyAlaValAlaSerPheLeuATGATGAsn145nGln11 152
      |||||||
      152 CCAGAGTGAAGCGAGTGTAGTTCTTGTGAGAAAGAAATATCAACAGAT 201
      |||||||
      152 eLysAlaTyrlIleSerMetHisSerTySerGlnHisIleValPhePro 169
      |||||||
      202 TAAAGCATATACATGATCATATCTATCTCCAGCATATATGTTTCAT 251
      |||||||
      169 yRserTyThrArgSerLysSerLysAspHisGluGluLeuSerLeuVal 185

```

```

|||||
253 ATTCTATACAGAAAGTAAAGCAAGACCATGAGAACTGCTCTAGTA 301
186 AAlaSerGluAlaValArgAlaIleGluTyrSerLysAsnThrArgTy 202
302 GCCAGTGAAGCAGTTGCTGCTATTGAGAAACACTAGTAAATAATCACGTA 351
202 rThrsHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyA 219
352 TACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGGAGTGGGG 401
219 sPASPTrpIleTyrAspLeuGlyIleLysTyrSer Phe..... 231
402 ACGATTGGATCTATGATTGGCATCAAAATATCGTTTACAATTGAACCT 451
232 .....ThrsAsp 235
452 CGAGATACGGGCAACATACGATTCTTGCCTGGAGCGATTACATCAAAAC 501
235 rProValGluLysLeuLeuProLeuSerLeuLys 246
|||||
502 CACCTGTAGAGAACGCTTTGCCGCTGCTCTAATA 536

seq_name: gb_est1:AL536068

seq_documentation_block:
LOCUS AL536068 575 bp mRNA linear EST 13-FEB-2001
DEFINITION AL536068 LTI_FL013_FBRn1 Homo sapiens cDNA clone CS0DF022YE21 3
prime, mRNA sequence.
ACCESSION AL536068
VERSION AL536068.1 GI:12799561
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 575)
AUTHORS L.M.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
Location/Qualifiers
1..575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF022YE21"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by life
technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
Maryland 20850, USA Fax : (1) 301 610 8371 Email :
filiang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 151 a 129 c 125 g 168 t 2 others
ORIGIN
alignment_scores:
Quality: 812.50 Length: 182
Ratio: 5.015 Gaps: 3
Percent Similarity: 89.011 Percent Identity: 87.363
alignment_block:

```

```

US-09-980-881-4 x AL536068/rev ..
Align seq 1/1 to reverse of: AL536068 from: 1 to: 575
84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysI 100
|||||
556 AATCGAATGTGGAGAAAGAACCGTTCTTTATGCCAACAATCATTCAT 507
100 eGlyTrpAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGlu 117
506 TGGACACAGACCTGAATAGAAACCTTGCTTCCAAACACTGGTGAGGAG 457
117 LysAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGlu 133
456 GTGCATCCAACTTCCTCATGCTCGGAAACCTACTGTGAGACTTATCCG 407
134 SerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnLeuA 150
406 TCAGAACCAAGAGTGAAGGCGATGGCTAGTTCTTGAGAGAAATATCAA 357
150 sngInIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 166
356 GCCAGATTAAAGCATACATCAGCATGATTCATATACCCAGCATATAGTG 307
167 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeu 183
306 TTTCATATTCTCTATACAGCAAGTAAAGCAAGACCATGAGAACTGTC 257
183 rLeuValAlaSerGluValAlaArgAlaIleGluLysThrSerLysAsnT 200
256 TCTAGTAGCCAGTGAACCACTGCTGCTATTGAGAAACACTAGTAAATA 207
200 hrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGly 216
206 CCAGGTATACATGCGCATGGCTCAGAAACCTATACCTAGCTCCGGA 157
217 GlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThr.. 232
156 GGTGGGACCATGTGATCTGATTTGGCATCAATTTTCGTTTACAAAT 107
233 .....S 233
106 TGAACCTTCGAGTTACGGGCACATACGATTCTTGCTGCCGGGCGTTTCC 57
233 eRAsn.ProProValGluLysLeuProLeuSerLeuLys 246
56 TCAACCCCGCCCTGAGAGAGCTTTGCCGCTGCTCTAATA 15

seq_name: gb_est1:AL536069

seq_documentation_block:
LOCUS AL536069 889 bp mRNA linear EST 13-FEB-2001
DEFINITION AL536069 LTI_FL013_FBRn1 Homo sapiens cDNA clone CS0DF022YE21 5
prime, mRNA sequence.
ACCESSION AL536069
VERSION AL536069.1 GI:12799562
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 889)
AUTHORS L.M.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..889
/organism="Homo sapiens"
/db_xref="taxon:9606"

```



```

|||||
426 AGAAGCTCTAGTGGCCAGGACGACAGTCTGCAATTGAAAGTATTA 377
197 eRLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
376 ATMAAACCCAGGACGACACGACGAGTGGCTCAGAAAGTTATATCTTA 327
214 AAlaProGlyGlyAlaAspAspTrpIleTyrAspLeuGlyIleLysTyrSe 230
326 GCTCCGTGGAGTTCAGACGATGATCTATGATTGGCATCAATATATTC 277
230 r.Phe..... 231
276 GTTTCACATTGAGCTCCGAGATACAGGACAGATACGAGATTCTGCTGCTG 227
232 .....ThSerAsnProProValGluLysLeuLeuProLeuSerLeuL 246
226 AGAGTACATCAAAACCCACTTGTGCAGAAAGCTTTGGCCCATCTCTTAA 177
246 s 246
176 A 176
seq_name: gb_est2:BF384322
seq_documentation_block:
LOCUS BF384322 692 bp mRNA linear EST 27-NOV-2000
DEFINITION 602046670F1 NCL_CGAP_119 Mus musculus cDNA clone IMAGE:4196129 5',
mRNA sequence.
ACCESSION BF384322
VERSION BF384322.1 GI:11365627
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 692)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.lnl.gov
Plate: L14M9532 row: 1 column: 18
High quality sequence stop: 689.
FEATURES
source
location/Qualifiers
1..692
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="4196129"
/lab_host="NCL_CGAP_119"
/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1; Note:
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCL_CGAP library."
BASE COUNT 211 a 156 c 139 g 186 t
ORIGIN
alignment_scores:
Quality: 676.50 Length: 174
Ratio: 4.340 Gaps: 2
Percent Similarity: 85.632 Percent Identity: 76.437
alignment_block:

```

```

US-09-980-881-4 x BF384322 ..
Align seg 1/1 to: BF384322 from: 1 to: 692
91 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAs 107
|||||
1 CGCTGTGCTCACAGAACGACCGCTCCGTGGGCACAGACCTGAACAGGAA 50
107 nPheAlaSerLysHisTrpCysGluGluGluIleValAspSerSerCys 124
51 CTTCGCTTCCAAACACTGCTGTGAGAAAGCTGCTCAAGTCTCTCTGCT 100
124 eRLuThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
|||||
101 CTGAACCTACTGTGACTTTATCTGACTGTGAGCAGAGTGAAGGCA 150
141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrLise 157
|||||
151 GTGGCTGACTTCTTGAGAAATATCGACACATTAAAGCTTACATCAG 200
157 rMetHisSerTyrSerGlnHisIleValProTyrSerTyrTrpArgS 174
201 TATGCATCATCTATCCCAACAAATATCTGTTCCCTATTCCTATTAACAGA 250
174 eRLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 190
|||||
251 GAAAAGCAAGCAGCAGCAAGAACTGTCTGTAGTGGCCAGCAACAGATT 300
191 ArgAlaIleGlyLysThrSerLysAsnThrArgTyrThrHisGlyHisG 207
|||||
301 CGTGCATTTGAAAGTATTAATTAACACACAGTACACACAGCGAGTGG 350
207 ySerGluThrLeuTyrLeuAlaProGlyGlyAlaAspAspTrpIleTyr 224
|||||
351 CTCGAAAGTTATATATCTAGCTCCGAGAGTCTCAGCATTTGATCTATG 400
224 sPLeuGlyLysTyrSer.Phe..... 231
|||||
401 ATTTGGCATCAAAATATTCGTTCAATTAAGCTCCGAGATACAGCAGA 450
232 .....ThSerAsnProProValGluLysL 240
451 TACGATTTCTGCTCCGAGAGATACACAAACCCACTGTGCAGAAAGC 500
240 eULeuProLeuSerLeuLys 246
|||||
501 TTGGCCGCCATCTCTAATAA 520
seq_name: gb_est1:A1255397
seq_documentation_block:
LOCUS A1255397 747 bp mRNA linear EST 12-NOV-1998
DEFINITION u194a04.x1 Sugano mouse liver mla Mus musculus cDNA clone
IMAGE:1690030 3' similar to TR:015114 Q15114 PREPRO-PLASMA
CARBOXYPEPTIDASE B.; mRNA sequence.
ACCESSION A1255397
VERSION A1255397.1 GI:3862922
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 747)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicinep

```


4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:974354

Possible reversed clone: similarity on wrong strand
Seq primer: custom primer used
High quality sequence stop: 471.
Location/Qualifiers

FEATURES

source

1..747
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone_image="IMAGE:1890030"
/clone_1lb="Sugano mouse liver ml1a"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: pME18S-F13; Site_1: DraIII
(CACTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor (TGTGGCCTACTG), digested
and cloned into distinct DraIII sites of the pME18S-F13
vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTTAACACTGCG and 3' end
primer CGACCTGCACCTCGACACA."

BASE COUNT 199 a 151 c 162 g 233 t 2 others
ORIGIN

alignment_scores:
Quality: 659.50 Length: 181
Ratio: 4.368 Gaps: 2
Percent Similarity: 83.425 Percent Identity: 72.376

alignment_block:
US-09-980-881-4 x AI255397/rev ..

Align seg 1/1 to reverse of: AI255397 from: 1 to: 747

```

84 AsnArgMetIrpArpLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
|||||
711 AATCGAATGTGGAGGAGACCGCTTGTCTCAACAGACCAACCGTGGNGT 662
|||||
100 eGlyThrAspLeuAsnArgAsnPheAlaSerLysHisIstPrcySgluIug 117
|||||
661 GGGACAGACAGCTGACACGAGACTTCTTCAACACTGGTGTGAGAAAG 612
|||||
117 LysAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGlu 133
|||||
611 GTTCGCGAAGTTCCTCTCTTAACCTACTGTGACTTATCTCTGAG 562
|||||
134 SerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAs 150
|||||
561 TCTGACCCGAGAGTGAAGGCACTGGCTGACTTCTTGAAGAAATATCGA 512
|||||
150 ngInIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValp 167
|||||
511 CCACATTAAGCTTACATCATGATGCACTATCTCCCAACAAATCTGT 462
|||||
167 hepProTyrSerTyrThrArgSerLysSerLysAspHisgluIugLeuSer 183
|||||
461 TTCCCATTCCTATATACAGAACAAAGCAGACCCGACAAAGACTGTCT 412
|||||
184 LeuValAlaSerGluAlaValAlaArgAlaIleGluLysThrSerLysAsn 200
|||||

```

```

411 CTAGTGGCCAGCGAAGCAGTTCGTGCAATTGAAGTATTAATAAACAC 362
200 rArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyG 217
|||||
361 CAGGTACACACAGCGAGCGTGGCTCAGAAAGTTTATCTAGCTCTGGAG 312
|||||
217 LysGlyAspAspTIPILerTyrAspLeuGlyIleLysTyrSer.Phe.... 231
|||||
311 GTTCTGACAGATTGGATCTATGATTTGGCAGATCAAAATATCGTTTACAAT 262
|||||
232 ..... This 233
261 GAGTCCGAGATACAGCGACAGATTCGCTTCTGCTGCTGAGATACAT 212
|||||
233 eRAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
|||||
211 CAACCCACATTTGTGACAGACGTTTGGCCGCGATCTTAATAA 171
|||||

```

seq_name: gb.est2:BI246210

seq_documentation_block:

LOCUS BI246210 713 bp mRNA linear EST 17-JUL-2001
DEFINITION 602958724F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5124700 5',
mRNA sequence.

ACCESSION BI246210

KEYWORDS BI246210.1 GI:14789700

SOURCE EST.

ORGANISM house mouse.

Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 713)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov
Plate: L1AM11304 row: 0 column: 05

High quality sequence stop: 686.

FEATURES

source

1..713
/organism="Mus musculus"
/strain="FVB/N"

/db_xref="taxon:10090"
/clone_image="IMAGE:5124700"

/clone_1lb="NCI_CGAP_L19"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."

BASE COUNT 212 a 168 c 145 g 188 t
ORIGIN

alignment_scores:
Quality: 656.50 Length: 175
Ratio: 4.406 Gaps: 3
Percent Similarity: 85.143 Percent Identity: 74.857

alignment_block:
US-09-980-881-4 x BI246210 ..

Align seg 1/1 to: BI246210 from: 1 to: 713

```

91 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp.LeuAsnArgA 107

```



```

seq_name: gb_est2:H61865

seq_documentation_block:
LOCUS       H61865               546 bp      mRNA          linear   EST_06-OCT-1995
DEFINITION  yf02b04.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:204079 5' similar to SP:A41204 A41204 CARBOXYPEPTIDASE B ; ,
mRNA sequence.
ACCESSION   H61865
VERSION     H61865.1  GI:1014697
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 546)
AUTHORS     Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            M., Hulthan,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rickio,L., Rohlfing,T., Soares,M., Tan,F., Trevaakis,E., Waterston
            R., Williamson,A., Wollmann,P. and Wilson,R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel.: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 973
            High quality sequence stops: 377
            This clone is available royalty-free through LNLN ; contact the
            IMAGE Consortium (infoimage.llnl.gov) for further information.
            Insert Length: 973 Std Error: 0.00
            Seq primer: M13Rpl
            High quality sequence stop: 377.

FEATURES             location/qualifiers
     source           1..546
                     /organism="Homo sapiens"
                     /db_xref="GDB:3773210"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:204079"
                     /clone_lib="Soares fetal liver spleen INFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
                    with a modified polylinker; Site.1: Pac I; Site.2: Eco RI.;
                    1st strand cDNA was primed with a Pac I - oligo(dT) primer
                    [5' AACGTGAAGATTATTAATAAGATCTTTTCTTTTTTTT 3'] ,
                    double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Pac I and cloned into the Pac I
                    and Eco RI sites of the modified pT73 vector. Library
                    went through one round of normalization. Library
                    constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT        148 a      112 c      121 g      159 t              6 others
ORIGIN

alignment_scores:
Quality:         631.00      Length:         154
Ratio:           4.780      Gaps:           4
Percent Similarity: 85.714      Percent Identity: 84.416

alignment_block:
US-09-980-881-4 x H61865      ..

Align seg 1/1 to: H61865 from: 1 to: 546

113 TTPCYGUGLUGLUALASerSerSerCySserGCUThrTyTCysG 129
|||||
13 TGCGTGAGAGAAGCTCATCCACTTCCGTCGCGAACAACCTACTGTGG 62

129 yfeutyrrProgluSerGluProGluValIySalavaAlaSerPheLeuA 146

```

```

63 ACTTATTCCTGAGTGCAACCGAAGACTGAAGCAGGTAGTTCTTGA   |||||
146 TGAATGAAATTCAAACCAGATTAAAGCATATGCATGCATCATCTCC   |||||
113 GAAGAATAATCAAACCAGATTAAAGCATATGCATGCATCATCTCC   |||||
163 GlnHisIleVal.PheProTySerTrpThrArgSerLeysSerLysAspHl 179
|||||
163 CAGCAATATVGTGGTTCCATATTCCTATTAACAAGAAAAGAACAGCA    |
179 sclugJuleUserLeuValAlaSerGlutAlaValAlaIleGluLysr 196
|||||
213 TGAGGAAGCTGCTCTTAGTACGCCAGTGAACAGTTCGCTATTTAGAAAA 262
196 hSeerLysAnThArgrYrnHiSGlyhisGLy.SergLuThriDeuty 212
263 CTAGTAAATAATCACAGGTATACACATGGCCATNGCNCNAGAAACCTTAA 312
212 rleuaIapO.GlyglYglyaspAPtRilEtyraspleungliyelys 228
313 CCTAGCTCCTGGAGGTGGGAGCATTTGATGTATTTGGGCATCAAA 362
229 TyrSer..... 230
363 TATTCGTTTACATTGACACTTCGAGATACGGGACATACGGGTTCTTGC 412
231 .....PheThr.SerasnProproValIGlutylserLeuproLen 243
413 TCACGAGAGGCTTACATTCMAACCCACTGTTCAGAGAGTTTTGNCGCTG 462
244 SerLeu 245
|||||
463 TCTCTA 468

seq_name: gb_est1:A1595707

seq_documentation_block:
LOCUS A1595707 756 bp mRNA linear EST 21-APR-1999
DEFINITION uj68d08.x1 Sugano mouse liver mla Mus musculus cDNA clone
IMAGE:1923103 3' similar to TR:015114 Q15114 PREPRO-PLASMA
CARBOXYPEPTIDASE B.; , mRNA sequence.
ACCESSION AF595707
VERSION AF595707.1 GI:4604755
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 756)
AUTHORS Marra,M., Hallier,L., Kucaba,T., Martin,J.J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,N., Bowers,Y., Riter,E., Kohm,S.V., Shih,T., Jackson,Y., Cardenas,W., McCann,R., Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School Of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel.: 314 286 1800
Fax: 314 286 1810
Email: mousetest@watsn.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (infoimage.llnl.gov) for further information.
MGI:961395
This clone was previously sequenced on the 5' end only, this new data is from the 3' end
Possible reversed clone: similarity on wrong strand
Seq primer: custom primer used
High quality sequence stop: 511.
Location/Qualifiers
source 1..756
```



```
17 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
|||||
239 TTGGATAGAAATTATTAAGTAGAGGCACTCTGATAGCTTACAAAAATCC 288
34 IsIleGlySerSerPheGluIleTyrProLeuTyrValLeuLysValSer 50
|||||
289 ACATTTGATCCCATTTGAGAACTACCCACTCATGTGTTAAAGGTTCTT 338
51 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 67
|||||
339 GGAAAAAGAACACAGCACCAAAAATGCCATATGGATTGACTGTGAATCCA 388
67 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 83
|||||
389 TGCCAGAGAAATGACCTCTCTGCTTCTGCTTGCTGCTGATAGGCCATA 438
83 ..... 83
439 TAACTCATTCATGGGATATAGGGCAATATACCAATCTCTGAGGCTT 488
84 .....AsnArgMe 86
489 GTGGATTTCATGTATGCCAGTGGTTAATGTGATGTATGACTACTC 538
|||||
86 tTyrArg.....LysAsnArgSerPheTyrAla 96
|||||
539 ATGGAACACAGAAATCGAATGTGGAGAACAGAACCGTTCTTCTATGCGA 588
96 sAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHis 112
|||||
589 ACAATCATTCATCGGACAGACCTGATATGGAATT.GCTTCCAAACAC 637
113 TrpCysGluGln.GlyAlaSerSerSerSer.CysSerGlu.ThrTyr.C 128
|||||
638 TGGTGTGAGGAGGATGATCGATCGATTCCTCATGTCTCGAAAAACCTACTT 687
128 ysgLyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPhe 144
|||||
688 GTGGACTTTATCTCTGAGTCAGAACAGAGTG.GAAGCAGTGGCTAGTTTC 736
145 LeuArg 146
|||||
737 TTGAG 742
```

)

This Page Blank (uspto)